A numerical comparison among some algorithms for unconstrained non-linear function minimization

Heuts, R.M.J.; Rens, P.J.

Publication date:
1972

EIT
34

R. M. J. Heuts and P. J. Rens

A numerical comparison among some algorithms for unconstrained non-linear function minimization

Research memorandum

TILBURG INSTITUTE OF ECONOMICS
DEPARTMENT OF ECONOMETRICS

A numerical comparison among some algorithms for unconstrained non-lineair function minimization

by R.M.J.Heuts and P.J.Rens

Preliminary and confidential January, 1972

SUMMARY.

For unconstrained function minimalization we have analysed experimentally four algorithms which are:

a. Davidon-Fletcher-Powell-Shanno (DFPS) algorithm [3b, 5, 13, 16, 17] with Hermite interpolation;
b. DFPS algorithm, Lagrange interpolation with equidistant points [8];
c. DFPS algorithm, Lagrange interpolation with points not necessarily equidistant;
d. Marquardt's algorithm [11].

The procedures a, c and d are given in this report, the procedure b can be found in Heuts R.M.J. and W.H.Vandaele [8].
The first three algorithms are specially suited for minimalization of general convex functions.
Marquardt's procedure is developed for object functions which are sums of squares.
The above four algorithms were tested for two functions and several initial estimates and d was also applied to a logistic growth curve which is frequently used in economics.

CONTENTS.

1. INTRODUCTION.

The methods described in chapters 2 and 3 are based on an algorithm of Davidon W.C. [3b] and extended by Fletcher R. and M.J.D.Powell [5] and generalised in a series of articles by Shanno D.F. and P.C. Kettler [16 , 17]. It is an iterative procedure to find the relative minimum of a non-linear function of several variables without constraints.
For line minimalization (step length procedure) we have used two interpolation procedures e.g. Hermite and Lagrange interpolation.
The method of Marquardt (chapter 4) uses the special structure of an object function, which is the sum of squares of non-linear functions of several variables.
A function often used in economics, the logistic growth curve, is examined with one of the above procedures.

2. QUASI NEWTON ALGORITHM WITH CUBIC (HERMITE) INTERPOLATION.

2.1.1 Method of computation.

The non-linear function $Q(x ; \Theta)$ where x is a vector of known variables and Θ is the parameter vector, is minimized as follows:

- Given the parameter vector $\Theta^{(k)}$ we compute the gradient vector

$$g^{(k)} = \left\{ \frac{\partial Q(x ; \Theta)}{\partial \Theta} \right\}_{\Theta=\Theta^{(k)}} , \text{ a p x 1 matrix,}$$

  where k is the k-th stage of the algorithm.

- Thereafter we determine the direction vector

$$s^{(k)} \equiv s^{(k)}(x ; \Theta) = -H^{(k)} g^{(k)}$$

  The determination of the matrix $H^{(k)}$ will be discussed below.

- Next we search along $-H^{(k)} g^{(k)}$ to find a scalar $\alpha_{(k)}$ such that $Q(x ; \Theta^{(k)} + \alpha_{(k)} s^{(k)})$ is a minimum. Fletcher R. and M.J.D.Powell have proved that $\alpha_{(k)}$ may be a positive scalar [5].

- The scalar $(\alpha_{(k)})_{min}$ is used to determine the steplength along the line $s^{(k)}$:
  $\sigma^{(k)} = (\alpha_{(k)})_{min}\, s^{(k)}$, after which we can compute the new vector of parameters

  $$\theta^{(k+1)} = \theta^{(k)} + \sigma^{(k)}.$$

- Set $z^{(k)} = g^{(k+1)} - g^{(k)}$

- Compute

$$H^{(k+1)} = H^{(k)} + t\frac{\sigma^{(k)}\sigma^{(k)'}}{\sigma^{(k)'}z^{(k)}} + \frac{[(1-t)\sigma^{(k)} - H^{(k)}z^{(k)}][(1-t)\sigma^{(k)} - H^{(k)}z^{(k)}]'}{[(1-t)\sigma^{(k)} - H^{(k)}z^{(k)}]'z^{(k)}}$$

  The use of the parameter t will be discussed later.

- The algorithm is repeated until a minimum is reached.
  In the first step of the iteration it is customary to set $H^{(o)} = I$ so that the first step is equivalent to a step of the steepest descent method.
  If $H^{(o)}$ is positive definite, then it can be proved that all subsequent $H^{(k)}$ are also positive definite [16].

2.1.2 Starting value for the parameter vector $\theta^{(o)}$.

For the computation of starting values $\theta^{(o)}$ we can take a number of arbitrary parameter vectors in the p-dimensional parameter space and compute for every vector in this space the function value of $Q(x\ ;\ \theta)$.
That vector for wich the object function is a minimal, will

be taken as a starting vector for the iterative procedure.

2.1.3 The determination of $\alpha_{(k)}$.

The scalar parameter $\alpha_{(k)}$ can be found by fitting a second or third degree polynomial in $\alpha_{(k)}$ to the Q-values and to determine the value of $\alpha_{(k)}$ for wich this polynomial is a minimum.

2.1.4 The determination of the matrix $H^{(k)}$.

Newton's method for minimizing a function $Q(x ; \Theta)$, where Θ is a p-vector, is to generate a sequence of points

$$\Theta^{(k+1)} = \Theta^{(k)} - \alpha_{(k)} [S^{(k)}]^{-1} g^{(k)}, \ldots \qquad (2.1)$$

where $g^{(k)} = \left[\frac{\partial Q(x ; \Theta)}{\partial \Theta}\right]_{\Theta=\Theta^{(k)}}$, $S^{(k)} = \left[\frac{\partial^2 Q(x ; \Theta)}{\partial \Theta_i \partial \Theta_j}\right]$,

the Hessian matrix of $Q(x ; \Theta)$ evaluated at $\Theta^{(k)}$, and $\alpha_{(k)}$ an appropriately chosen scalar.
Quasi-Newton methods use an initial approximation and generate an approximation $H^{(k)}$ to $[S^{(k)}]^{-1}$ at each step rather than performing the computational work of evaluating and inverting $S^{(k)}$, such as by Hartley's method [7].
The sequence (2.1) then becomes

$$\Theta^{(k+1)} = \Theta^{(k)} - \alpha_{(k)} H^{(k)} g^{(k)}. \qquad (2.2)$$

The determination of $\alpha_{(k)}$ is discussed in (2.1.3).

Some well-known techniques of this type are the Fletcher-Powell modification of Davidon's method [5], Broyden methods [2, 3], the Barnes-Rosen method [1, 15] and Goldfarb's method [6]. The Fletcher-Powell technique guarantees that the matrix $H^{(k)}$ will always be positive definite.
Shanno's method [16] develops a family of matrices $H^{(k)}$ as a function of a scalar parameter t, and it can be shown that both the Fletcher-Powell and Barnes-Rosen matrices are special cases of this parametric family.
The technique for generating a series of approximations $H^{(k)}$ to the inverse of the Hessian at the points $\Theta^{(k)}$ can be described as follows:
Assume $Q(x ; \Theta)$ is a positive definite quadratic form, $H^{(k)}$ the current approximation to the inverse of the Hessian, and $S^{(k)}$ the approximation to the Hessian.
Assume $z^{(k)} = g^{(k+1)} - g^{(k)}$. If $S^{(k)}$ is a constant matrix, we have $g^{(k)} = \nabla Q = x + S\,\Theta^{(k)}$, and when $Q(x ; \Theta)$ can be approximated by the standard quadratic form in p-dimensions

$$Q(x ; \Theta) = Q_o + \sum_{i=1}^{p} \Theta_i x_i + \tfrac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} S_{ij} \Theta_i \Theta_j$$

then

$$z^{(k)} = S (\Theta^{(k+1)} - \Theta^{(k)}) \equiv S \sigma^{(k)}. \qquad (2.3)$$

Multiplying (2.3) by $H^{(k)}$, we obtain

$$H^{(k)} z^{(k)} = \sigma^{(k)}. \qquad (2.4)$$

Since (2.4) in general will not be satisfied, assume the error lies in $H^{(k)}$. The matrix $H^{(k)}$ has to a certain extent the properties of the inverse of the S-matrix. We then modify $H^{(k)}$ by introducing a matrix of additive corrections $D^{(k)}$ such that

$$(H^{(k)} + D^{(k)}) z^{(k)} = \sigma^{(k)} \qquad (2.5)$$

or

$$D^{(k)} z^{(k)} = \sigma^{(k)} - H^{(k)} z^{(k)} \qquad (2.6)$$

To restrict the choice of t, Shanno [16] has proved the following:

If $H^{(k)}$ is positive definite, $\forall\ t > \frac{\alpha_{(k)} - 1}{\alpha_{(k)}}$, also $H^{(k+1)}$ is positive definite.

For $t > \frac{\alpha_{(k)} - 1}{\alpha_{(k)}}$, $H^{(k+1)}$ is positive definite, hence at no finite step does the smallest eigenvalue of $H^{(k+1)}$ ever become zero. However, it is possible that if λ_1 is the smallest eigenvalue of $H^{(k+1)}$, $\lim_{k\to\infty} \lambda_1 = 0$.

This can only be due to computer rounding errors. In this case the iterative technique will degenerate as $k \to \infty$.

To attempt to alleviate this difficulty, we may at each step choose t in such a way as to maximize the smallest eigenvalue of $H^{(k+1)}$.

This is accomplished be choosing t to maximize $v' H^{(k+1)} v$ for any arbitrary vector v.

We shall show this as follows:

Say the matrix $H^{(k+1)}$ is positive definite, then maximizing the quadratic form $v' H^{(k+1)} v$ means maximizing $\sum_j \lambda_j v_j^2$,

where the λ_j are positive charateristic roots of the matrix $H^{(k+1)}$. Now we can say that maximizing $\sum_j \lambda_j v_j^2$ means also maximizing λ_1, the smallest charateristic root of the matrix $H^{(k+1)}$.

Shanno [16] has proved the following interesting theorem:

Let v be an arbitrary vector. Then $v' H^{(k+1)} v$ is a non-decreasing function of t.

Some well-known techniques of this type are the Fletcher-Powell modification of Davidon's method [5], Broyden methods [2, 3], the Barnes-Rosen method [1, 15] and Goldfarb's method [6]. The Fletcher-Powell technique guarantees that the matrix $H^{(k)}$ will always be positive definite.

Shanno's method [16] develops a family of matrices $H^{(k)}$ as a function of a scalar parameter t, and it can be shown that both the Fletcher-Powell and Barnes-Rosen matrices are special cases of this parametric family.

The technique for generating a series of approximations $H^{(k)}$ to the inverse of the Hessian at the points $\Theta^{(k)}$ can be described as follows:

Assume $Q(x ; \Theta)$ is a positive definite quadratic form, $H^{(k)}$ the current approximation to the inverse of the Hessian, and $S^{(k)}$ the approximation to the Hessian.

Assume $z^{(k)} = g^{(k+1)} - g^{(k)}$. If $S^{(k)}$ is a constant matrix, we have $g^{(k)} = \nabla Q = x + S\,\Theta^{(k)}$, and when $Q(x ; \Theta)$ can be approximated by the standard quadratic form in p-dimensions

$$Q(x ; \Theta) = Q_o + \sum_{i=1}^{p} \Theta_i x_i + \tfrac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} S_{ij} \Theta_i \Theta_j$$

then

$$z^{(k)} = S\,(\Theta^{(k+1)} - \Theta^{(k)}) \equiv S\,\sigma^{(k)}. \qquad (2.3)$$

Multiplying (2.3) by $H^{(k)}$, we obtain

$$H^{(k)}\, z^{(k)} = \sigma^{(k)}. \qquad (2.4)$$

Since (2.4) in general will not be satisfied, assume the error lies in $H^{(k)}$. The matrix $H^{(k)}$ has to a certain extent the properties of the inverse of the S-matrix. We then modify $H^{(k)}$ by introducing a matrix of additive corrections $D^{(k)}$ such that

$$(H^{(k)} + D^{(k)})\, z^{(k)} = \sigma^{(k)} \qquad (2.5)$$

or

$$D^{(k)}\ z^{(k)} = \sigma^{(k)} - H^{(k)}\ z^{(k)} \qquad (2.6)$$

To restrict the choice of t, Shanno [16] has proved the following:
If $H^{(k)}$ is positive definite, $\forall\ t > \frac{\alpha_{(k)} - 1}{\alpha_{(k)}}$, also $H^{(k+1)}$ is positive definite.

For $t > \frac{\alpha_{(k)} - 1}{\alpha_{(k)}}$, $H^{(k+1)}$ is positive definite, hence at no finite step does the smallest eigenvalue of $H^{(k+1)}$ ever become zero. However, it is possible that if λ_1 is the smallest eigenvalue of $H^{(k+1)}$, $\lim_{k\to\infty} \lambda_1 = 0$.

This can only be due to computer rounding errors. In this case the iterative technique will degenerate as $k \to \infty$.
To attempt to alleviate this difficulty, we may at each step choose t in such a way as to maximize the smallest eigenvalue of $H^{(k+1)}$.
This is accomplished be choosing t to maximize $v'\ H^{(k+1)}\ v$ for any arbitrary vector v.
We shall show this as follows:
Say the matrix $H^{(k+1)}$ is positive definite, then maximizing the quadratic form $v'\ H^{(k+1)}\ v$ means maximizing $\sum_j \lambda_j\ v_j^2$,

where the λ_j are positive charateristic roots of the matrix $H^{(k+1)}$. Now we can say that maximizing $\sum_j \lambda_j\ v_j^2$ means also

maximizing λ_1, the smallest charateristic root of the matrix $H^{(k+1)}$.

Shanno [16] has proved the following interesting theorem:
Let v be an arbitrary vector. Then $v'\ H^{(k+1)}\ v$ is a non-decreasing function of t.

It can be shown that the condition of $H^{(k+1)}$ improves monotonically with t. This necessitates finding a closed form representation of $H^{(k+1)}$ for $t = \infty$.
It can be proved that this representation is as follows [16]:
Let $H^{(k+1)}$ be defined by (2.10), and let

$$r = \frac{\sigma^{(k)'} z^{(k)}}{\sigma^{(k)'} z^{(k)} + z^{(k)'} H^{(k)} z^{(k)}}$$

then

$$\lim_{t\to\infty} H^{(k+1)} = H^{(k)} + \frac{(\sigma^{(k)} - r H^{(k)} z^{(k)})(\sigma^{(k)} - r H^{(k)} z^{(k)})'}{(\sigma^{(k)} - r H^{(k)} z^{(k)})' z^{(k)}} +$$

$$+ (r-1) \frac{H^{(k)} z^{(k)} z^{(k)'} H^{(k)}}{z^{(k)'} H^{(k)} z^{(k)}}$$

If we let $H^{(k+1)} = H^{(k)} + D^{(k)}$, and $D^{(k)}$ is chosen to satisfy (2.6), we then have

$$H^{(k)} z^{(k)} = \sigma^{(k)}. \qquad (2.7)$$

Fletcher and Powell [5] have proved that when $Q(x ; \Theta)$ is a positive definite quadratic form, and at each step (2.7) is satisfied, then the minimum of $Q(x ; \Theta)$ will be reached in at most p iterations.
For functions which are not quadratic, but strict convex, the convergence can also be proved [13].
To determine $D^{(k)}$ we can write (2.6) in the form

$$D^{(k)} z^{(k)} = t \sigma^{k} + (1-t) \sigma^{(k)} - H^{(k)} z^{k} \qquad (2.8)$$

which by multiplying with the scalars

$$\sigma^{(k)\prime}{}_{z}{}^{(k)} \quad \text{and} \quad ((1-t)\sigma^{(k)}-H^{(k)}{}_{z}{}^{(k)})'\ {}_{z}{}^{(k)}$$

becomes

$$D^{(k)}{}_{z}{}^{(k)}=t\frac{\sigma^{(k)}\sigma^{(k)\prime}{}_{z}{}^{(k)}}{\sigma^{(k)\prime}{}_{z}{}^{(k)}}+\frac{[(1-t)\sigma^{(k)}-H^{(k)}{}_{z}{}^{(k)}][(1-t)\sigma^{(k)}-H^{(k)}{}_{z}{}^{(k)}]'_{z}{}^{(k)}}{[(1-t)\sigma^{(k)}-H^{(k)}{}_{z}{}^{(k)}]'\ {}_{z}{}^{(k)}}.$$

So

$$D^{(k)}=t\frac{\sigma^{(k)}\sigma^{(k)\prime}}{\sigma^{(k)}{}_{z}{}^{(k)}}+\frac{((1-t)\sigma^{(k)}-H^{(k)}{}_{z}{}^{(k)})((1-t)\sigma^{(k)}-H^{(k)}{}_{z}{}^{(k)})'}{((1-t)\sigma^{(k)}-H^{(k)}{}_{z}{}^{(k)})'\ {}_{z}{}^{(k)}}. \tag{2.9}$$

Then t = 0 is the Barnes-Rosen choice, and t = 1 the Fletcher-Powell choice.
So we have

$$H^{(k+1)}=H^{(k)}+t\frac{\sigma^{(k)}\sigma^{(k)\prime}}{\sigma^{(k)\prime}{}_{z}{}^{(k)}}+$$

$$\frac{[(1-t)\sigma^{(k)}-H^{(k)}{}_{z}{}^{(k)}][(1-t)\sigma^{(k)}-H^{(k)}{}_{z}{}^{(k)}]'}{[(1-t)\sigma^{(k)}-H^{k}{}_{z}{}^{(k)}]'\ {}_{z}{}^{(k)}} \tag{2.10}$$

Powell [13] has proved under general conditions that in the variable metric algorithm

$$\| \theta^{(k+1)} - \theta^{*} \| \leq M \| \theta^{(k)} - \theta^{*} \|,$$

where θ^{*} is the value of θ minimizing $Q(x ; \theta)$ and $\| \ \|$ is the Euclidean vector norm.

This suggests using $\| \sigma^{(k)} \| = \| \alpha_{(k)} s^{(k)} \| = \| -\alpha_{(k)} H^{(k)} g^{(k)} \|$ as an estimate of $\| \theta^{(k)} - \theta^{*} \|$ and t is chosen so that $\| s^{(k+1)} \| = \| \sigma^{(k)} \|$. This version is called the constant norm version.

2.2 Line searches with a cubic interpolation formula (Hermite interpolation).

The cubic interpolation technique devised by Davidon [3b] is used to locate $\alpha_{(k)}$ at each step after which two points are found at which

$$\frac{d\,Q\,(x\,;\,\Theta)}{d\,\alpha_{(k)}} < 0 \quad \text{and} \quad \frac{d\,Q\,(x\,;\,\Theta)}{d\,\alpha_{(k)}} > 0.$$

A necessary condition for $\alpha_{(k)}$ to minimize $Q(x\,;\,\Theta)$ along $Q(x\,;\,\Theta^{(k)} + \alpha_{(k)}\,s^{(k)})$ is that

$$\frac{d\,Q\,(u)}{d\,\alpha_{(k)}} = \frac{d\,Q\,(u)}{d\,u} \cdot \frac{d\,u}{d\,\alpha_{(k)}} = g^{(k+1)'}\,s^{(k)} =$$

$$= -\,g^{(k+1)'}\,H^{(k)}\,g^{(k)} = 0\,.$$

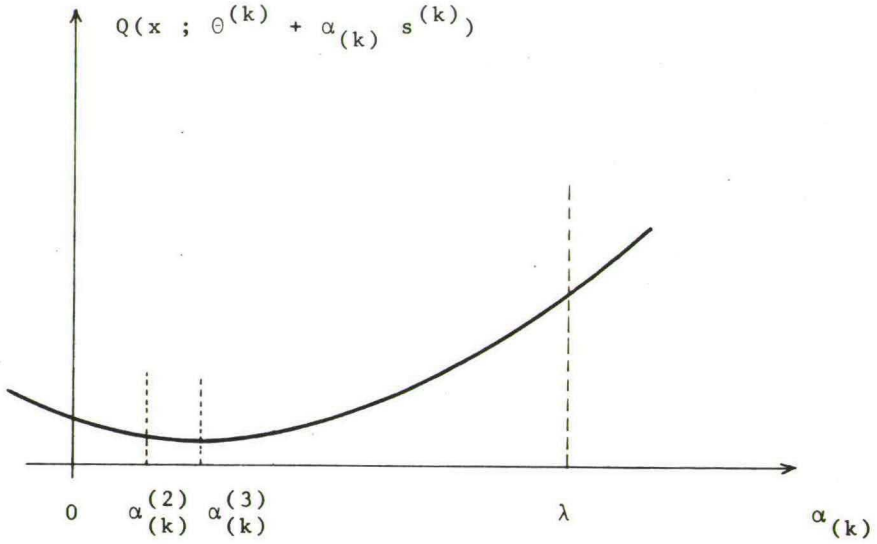
As long as $-\,g^{(k+1)'}\,H^{(k)}\,g^{(k)} < 0$ the function $Q(x\,;\,\Theta)$ is still decreasing and the steplength $\alpha_{(k)}$ must be increased. But after a while $-\,g^{(k+1)'}\,H^{(k)}\,g^{(k)} > 0$ and the function $Q(x\,;\,\Theta)$ is increasing again. So somewhere in between $-\,g^{(k+1)'}\,H^{(k)}\,g^{(k)} = 0$, and we are searching the $\alpha_{(k)}$ which satisfies this condition.

Davidon's interpolation method is nothing more than Hermite interpolation, which uses the value of the function and its derivative at two points to fit a cubic approximation. For a reference to Hermite interpolation we suggest any book on approximation theory, such as Macon N. [10] and Weeg G.P. and G.B.Reed [19].

If the two points are $\alpha^{1}_{(k)} = 0$ and $\alpha^{2}_{(k)} = \lambda$ as in Davidon's paper (see also appendix II in Heuts R.M.J. and W.H.Vandaele [8]) the resulting system of equations has the solution documented by Davidon. If, however, the two points are arbitrarily, the resulting system of equations must be solved.

To illustrate the more general method of Hermite interpolation, assume the following case (see figure 2.1):

figure 2.1



Using

$$Q\left[x\ ;\ \Theta^{(k)} + \alpha_{(k)}\ s^{(k)}\right]\Big]_{\alpha_{(k)}} = 0 \qquad \text{and}$$

$$Q\left[x\ ;\ \Theta^{(k)} + \alpha_{(k)}\ s^{(k)}\right]\Big]_{\alpha_{(k)}} = \lambda$$

and

$$\left.\frac{d\,Q\left[x\,;\,\Theta^{(k)}+\alpha_{(k)}\,s^{(k)}\right]}{d\,\alpha_{(k)}}\right]_{\alpha_{(k)}=0} \quad \text{and}$$

$$\left.\frac{d\,Q\left[x\,;\,\Theta^{(k)}+\alpha_{(k)}\,s^{(k)}\right]}{d\,\alpha_{(k)}}\right]_{\alpha_{(k)}=\lambda} ,$$

Davidon's and the general Hermite interpolation method each obtain $\alpha_{(k)}^{(2)}$ as the new estimate of the minimum. Davidon does not make it clear what he would do next i.e. whether to accept the point or proceed with a new estimate. As convergence criterion we will in the general Hermite method, not accept $\alpha_{(k)}^{(2)}$ unless either

$$\left.\frac{d\,Q\left[x\,;\,\Theta^{(k)}+\alpha_{(k)}\,s^{(k)}\right]}{d\,\alpha_{(k)}}\right]_{\alpha_{(k)}=\alpha_{(k)}^{(2)}} > 0$$

or

$$\left|\left.\frac{d\,Q\left[x\,;\,\Theta^{(k)}+\alpha_{(k)}\,s^{(k)}\right]}{d\,\alpha_{(k)}}\right]_{\alpha_{(k)}=\alpha_{(k)}^{(2)}}\right| < \delta$$

but in many cases we wish to perform another iteration,

using $\left.Q\left[x\,;\,\Theta^{(k)}+\alpha_{(k)}\,s^{(k)}\right]\right]_{\alpha_{(k)}=\alpha_{(k)}^{(2)}}$,

$$\left.Q\left[x\,;\,\Theta^{(k)}+\alpha_{(k)}\,s^{(k)}\right]\right]_{\alpha_{(k)}=\lambda} ,$$

$$\left.\frac{d\,Q\left[x\,;\,\Theta^{(k)}+\alpha_{(k)}\,s^{(k)}\right]}{d\,\alpha_{(k)}}\right|_{\alpha_{(k)}=\alpha^{(2)}_{(k)}} \quad \text{and}$$

$$\left.\frac{d\,Q\left[x\,;\,\Theta^{(k)}+\alpha_{(k)}\,s^{(k)}\right]}{d\,\alpha_{(k)}}\right|_{\alpha_{(k)}=\lambda} \quad , \text{ as}$$

these values bracket the minimum more closely than $(0,\lambda)$.

Since $\alpha^{(2)}_{(k)} \neq 0$, however, Davidon's formula's do not hold. For this reason we used the more general Hermite approach which can be described as follows:

We now define

$$Q\left[x\,;\,\Theta^{(k)}+\alpha^{(i)}_{(k)}\,s^{(k)}\right] \equiv Q\,(\alpha^{(i)}_{(k)})$$

and

$$\frac{d\,Q\left[x\,;\,\Theta^{(k)}+\alpha^{(i)}_{(k)}\,s^{(k)}\right]}{d\,\alpha^{(i)}_{(k)}} \equiv \frac{d\,Q\,(\alpha^{(i)}_{(k)})}{d\,\alpha^{(i)}_{(k)}}\,.$$

Given the functions:

$$Q\left[\alpha^{(i)}_{(k)}\right] = a\left[\alpha^{(i)}_{(k)}\right]^3 + b\left[\alpha^{(i)}_{(k)}\right]^2 + c\,\alpha^{(i)}_{(k)} + d$$

$$Q\left[\alpha^{(i-1)}_{(k)}\right] = a\left[\alpha^{(i-1)}_{(k)}\right]^3 + b\left[\alpha^{(i-1)}_{(k)}\right]^2 + c\,\alpha^{(i-1)}_{(k)} + d$$

$$\frac{d\,Q\left[\alpha^{(i)}_{(k)}\right]}{d\,\alpha^{(i)}_{(k)}} = 3\,a\left[\alpha^{(i)}_{(k)}\right]^2 + 2\,b\,\alpha^{(i)}_{(k)} + c \qquad \text{and}$$

$$\frac{d\,Q\left[\alpha_{(k)}^{(i-1)}\right]}{d\,\alpha_{(k)}^{(i-1)}} = 3\,a\left[\alpha_{(k)}^{(i-1)}\right]^2 + 2\,b\,\alpha_{(k)}^{(i-1)} + c,$$

then the coefficients a, b, c and d can be found as follows:

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \left[\alpha_{(k)}^{(i)}\right]^3 & \left[\alpha_{(k)}^{(i)}\right]^2 & \alpha_{(k)}^{(i)} & 1 \\ \left[\alpha_{(k)}^{(i-1)}\right]^3 & \left[\alpha_{(k)}^{(i-1)}\right]^2 & \alpha_{(k)}^{(i-1)} & 1 \\ 3\left[\alpha_{(k)}^{(i)}\right]^2 & 2\,\alpha_{(k)}^{(i)} & 1 & 0 \\ 3\left[\alpha_{(k)}^{(i-1)}\right]^2 & 2\,\alpha_{(k)}^{(i-1)} & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} Q\left[\alpha_{(k)}^{(i)}\right] \\ Q\left[\alpha_{(k)}^{(i-1)}\right] \\ \dfrac{d\,Q\,(\alpha_{(k)}^{(i)})}{d\,\alpha_{(k)}^{(i)}} \\ \dfrac{d\,Q\,(\alpha_{(k)}^{(i-1)})}{d\,\alpha_{(k)}^{(i-1)}} \end{bmatrix}$$

The values $\alpha_{(k)}^{(i)}$, $\alpha_{(k)}^{(i-1)}$, $Q\left[\alpha_{(k)}^{(i)}\right]$, $Q\left[\alpha_{(k)}^{(i-1)}\right]$, $\dfrac{d\,Q\left[\alpha_{(k)}^{(i)}\right]}{d\,\alpha_{(k)}^{(i)}}$ and $\dfrac{d\,Q\left[\alpha_{(k)}^{(i-1)}\right]}{d\,\alpha_{(k)}^{(i-1)}}$ are known and by matrix inversion we can find the values for a, b, c and d to solve the quadratic form:

$$\frac{d\,Q\,(\alpha_{(k)})}{d\,\alpha_{(k)}} = 3\,a\,(\alpha_{(k)})^2 + 2\,b\,\alpha_{(k)} + c = 0.$$

That positive valued root is taken for which $Q\,(\alpha_{(k)})$ is a minimum.

3. QUASI NEWTON ALGORITHM WITH QUADRATIC (LAGRANGE) INTERPOLATION.

3.1 See 2.1

3.2 Line searches with a quadratic interpolation formula (Lagrange interpolation).

For calculating the optimal steplength we make use of the following two theorems; a proof of them can be found in Macon N. [10].

Theorem 3.1

Let there be given n+1 ordered pairs of numbers (x_i , y_i), $i=0,1,\ldots,n$ where the x_i are distinct. Then there exists a unique polynomial

$$y^{(n)}(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_n$$

of degree n which satisfies the n+1 equations

$$y(x_i) = y_i \ , \ i=0,1,\ldots,n \ .$$

Theorem 3.2

The polynomial $y = y(x)$ of degree n which assumes the n+1 values y_i for n+1 distinct arguments x_i $(i=0,1,\ldots,n)$, can be written in the form

$$y^{(n)}(x) = \ell_0(x)\, y_0 + \ell_1(x)\, y_1 + \ldots + \ell_n(x)\, y_n$$

where the coefficients $\ell_i(x)$ are polynomials of degree n given by

$$\ell_i(x) = \frac{(x-x_0)(x-x_1)\ldots(x-x_{i-1})(x-x_{i+1})\ldots(x-x_n)}{(x_i-x_0)(x_i-x_1)\ldots(x_i-x_{i-1})(x_i-x_{i+1})\ldots(x_i-x_n)} .$$

It is not necessary for the arguments to be equidistant. Suppose we fit a second degree polynomial to three points. Then the polynomial has the following from

$$y^{(2)}(x) = \sum_{i=0}^{2} y_i \ell_i(x) =$$

$$= y_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + y_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}$$

$$+ y_2 \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} .$$

A minimum of this function is reached for

$$x_{min} = \frac{\frac{(x_1+x_2)\,y_0}{(x_0-x_1)(x_0-x_2)} + \frac{(x_0+x_2)\,y_1}{(x_1-x_0)(x_1-x_2)} + \frac{(x_0+x_1)\,y_2}{(x_2-x_0)(x_2-x_1)}}{\frac{2\,y_0}{(x_0-x_1)(x_0-x_2)} + \frac{2\,y_1}{(x_1-x_0)(x_1-x_2)} + \frac{2\,y_2}{(x_2-x_0)(x_2-x_1)}}$$

if the function is convex.
Transformed to the original variables (see 2.1) we have:

$$x_i \equiv \alpha_{(k)}^{(i)}$$

$$y_i \equiv Q\left\{x\ ;\ \Theta^{(k)} + \alpha_{(k)}^{(i)}\ s^{(k)}\right\} \equiv Q_{(k)}^{(i)}$$

and $y^{(2)}(x)$ will be defined as $P^{(2)}(x)$ with minimum $\alpha_{(k)}^{*}$. Entering the interpolation procedure we have at our disposal an $\alpha_{(k)}^{(1)}$ with derivative of the function

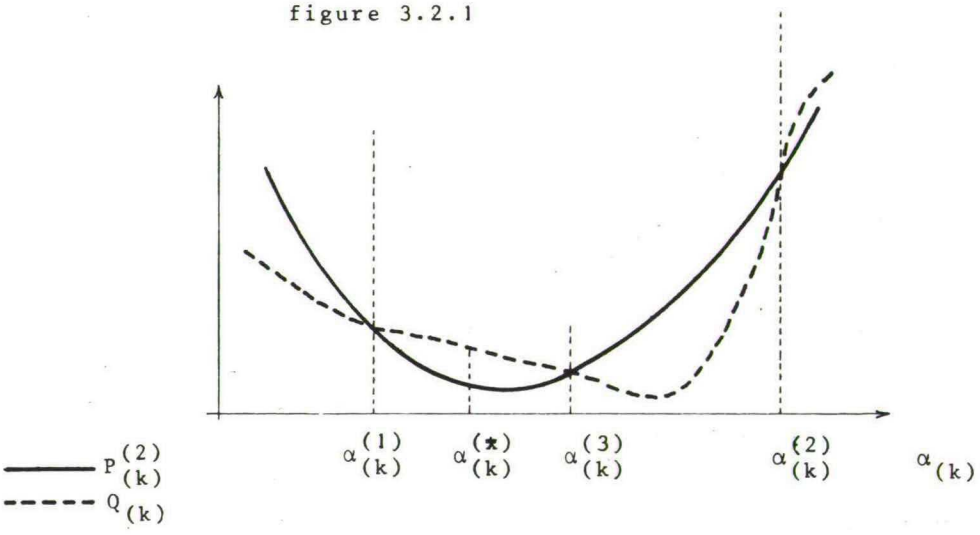
$\frac{d\ Q_{(k)}^{(1)}}{d\ \alpha_{(k)}^{(1)}} < 0$ and an $\alpha_{(k)}^{(2)}$ with derivative $\frac{d\ Q_{(k)}^{(2)}}{d\ \alpha_{(k)}^{(2)}} > 0$.

Then we choose $\alpha_{(k)}^{(3)}$ in the midth of the two α-values (Halving procedure). Concerning the convexity criterion there are two situations:

3.2.1 The convexity criterion is fullfilled.

3.2.1.1 Case 1.

figure 3.2.1



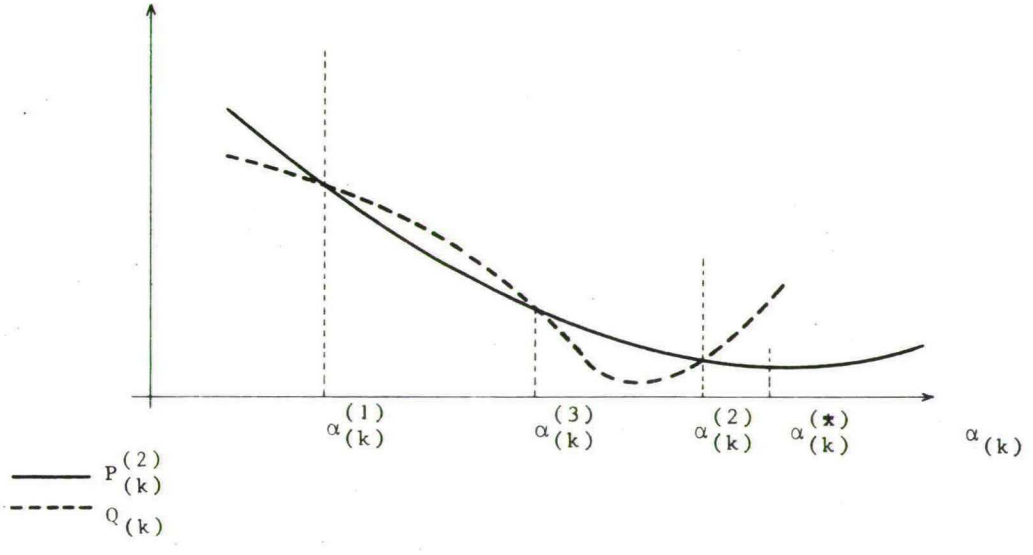
If $Q^{(3)}_{(k)} < Q^{(1)}_{(k)}$ and $Q^{(3)}_{(k)} < Q^{(2)}_{(k)}$ the quadratic interpolation formula can be repeated with the points $\alpha^{(*)}_{(k)}$, $\alpha^{(3)}_{(k)}$ and $\alpha^{(2)}_{(k)}$ (see figure 3.2.1) if

$$\frac{d\ Q^{(*)}_{(k)}}{d\ \alpha^{(*)}_{(k)}} < 0 \quad \text{and with } \alpha^{(1)}_{(k)} \text{ , } \alpha^{(*)}_{(k)} \text{ and } \alpha^{(3)}_{(k)} \text{ if } \frac{d\ Q^{(*)}_{(k)}}{d\ \alpha^{(*)}_{(k)}} > 0.$$

This is continued until two line coordinates b and d are obtained such that the corresponding points $\Theta^{(k)} + b\ s^{(k)}$ and $\Theta^{(k)} + d\ s^{(k)}$ differ in each component by less than an absolute accuracy $\varepsilon=0.01$.
Then the point with the smallest function value is delivered as an approximation of the line minimum.

3.2.1.2 Case 2.

figure 3.2.2



If $Q^{(2)}_{(k)} < Q^{(3)}_{(k)} < Q^{(1)}_{(k)}$ and $\frac{d\, Q^{(3)}_{(k)}}{d\, \alpha^{(3)}_{(k)}} < 0$ (see figure 3.2.2)

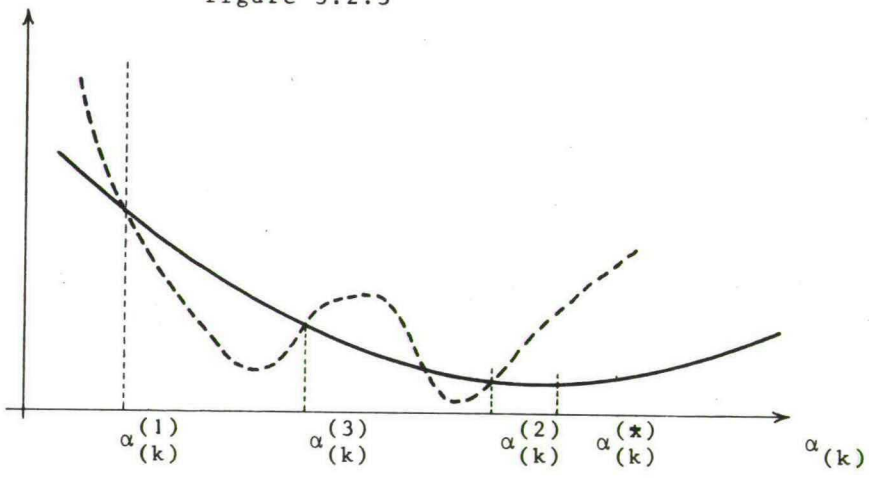
then we discard $\alpha^{(1)}_{(k)}$ and restart the halving procedure at the interval $\left[\alpha^{(3)}_{(k)} , \alpha^{(2)}_{(k)}\right]$ untill the situation of case 1 occurs.

3.2.1.3 Case 3.

If $Q^{(2)}_{(k)} < Q^{(3)}_{(k)} < Q^{(1)}_{(k)}$ and $\frac{d\, Q^{(3)}_{(k)}}{d\, \alpha^{(3)}_{(k)}} > 0$ (see figure 3.2.3)

then we know that there are at least two minima and we start the halving procedure for both sub-intervals until the situation of case 1 is satisfied. So two minima are calculated

figure 3.2.3



and we continue with the $\alpha_{(k)}$-value with the smallest function value.

If in the case of at least two minima $Q^{(2)}_{(k)} < Q^{(3)}_{(k)} < Q^{(1)}_{(k)}$ and $\frac{d\, Q^{(3)}_{(k)}}{d\, \alpha^{(3)}_{(k)}} < 0$ we proceed as in case 2, section 3.2.1.2..

3.2.1.4 Case 4.

If $Q^{(1)}_{(k)} < Q^{(3)}_{(k)} < Q^{(2)}_{(k)}$ and $\frac{d\, Q^{(3)}_{(k)}}{d\, \alpha^{(3)}_{(k)}} > 0$ we start the halving procedure at the interval $\left[\alpha^{(1)}_{(k)}\ ,\ \alpha^{(3)}_{(k)}\right]$.

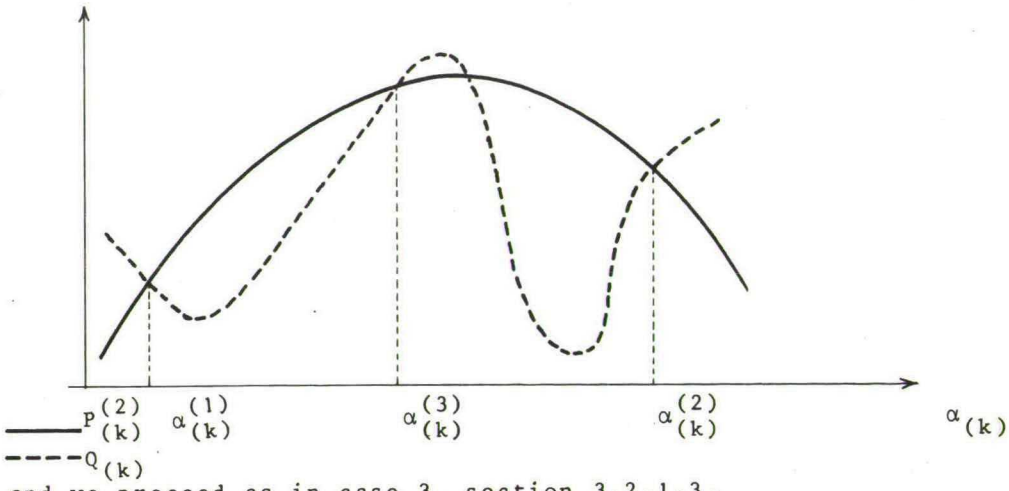
If $Q^{(1)}_{(k)} < Q^{(3)}_{(k)} < Q^{(2)}_{(k)}$ and $\frac{d\, Q^{(3)}_{(k)}}{d\, \alpha^{(3)}_{(k)}} < 0$ we proceed as in case 3, section 3.2.1.3.

3.2.2 The convexity criterion is not fullfilled.

3.2.2.1 Case 5.

If $Q^{(1)}_{(k)} < Q^{(3)}_{(k)}$ and $Q^{(2)}_{(k)} < Q^{(3)}_{(k)}$ then we are sure that there are at least two minima (see figure 3.2.4)

figure 3.2.4



and we proceed as in case 3, section 3.2.1.3.

3.2.2.2 Case 6.

If $Q^{(1)}_{(k)} < Q^{(3)}_{(k)} < Q^{(2)}_{(k)}$ and $\frac{d\,Q^{(3)}_{(k)}}{d\,\alpha^{(3)}_{(k)}} > 0$ we start the halving procedure at the interval $\left[\alpha^{(1)}_{(k)}\ ,\ \alpha^{(3)}_{(k)}\right]$.

If $Q^{(1)}_{(k)} < Q^{(3)}_{(k)} < Q^{(2)}_{(k)}$ and $\frac{d\,Q^{(3)}_{(k)}}{d\,\alpha^{(3)}_{(k)}} < 0$ we search two intervals as in case 3, section 3.2.1.3.

3.2.2.3 Case 7.

If $Q^{(2)}_{(k)} < Q^{(3)}_{(k)} < Q^{(1)}_{(k)}$ and $\frac{d\,Q^{(3)}_{(k)}}{d\,\alpha^{(3)}_{(k)}} > 0$ we proceed as in case 3.

If $Q^{(2)}_{(k)} < Q^{(3)}_{(k)} < Q^{(1)}_{(k)}$ and $\frac{d\,Q^{(3)}_{(k)}}{d\,\alpha^{(3)}_{(k)}} < 0$ we start the halving procedure on the interval $\left[\alpha^{(3)}_{(k)}\ ,\ \alpha^{(2)}_{(k)}\right]$.

4. THE ALGORITHM OF MARQUARDT.

The least squares method for non-linear regression based on the expansion of the model in a Taylor series through the first derivatives (the unmodified Gauss-Newton method) may frequently fail because of divergence and oscillation of the successive iterates, while the gradient or steepest descent methods often fail because of slow convergence after rapid initial progress on the first few iterations.
The Marquardt algorithm is, in effect, an interpolation between the Taylor series expansion and the steepest descent method. Marquardt found out that the angle α between the steepest descent vector and the Taylor series vector is for a variety of problems in the range $80^{o} < \alpha < 90^{o}$.
The algorithm is based on least squares of which the basic idea is as follows.
Suppose we wish to minimize $Q(x, \Theta)$ as a function of Θ where

$$Q(x, \Theta) = \sum_{i=1}^{n} [y_i - f(x_i, \Theta)]^2$$

using the following regression model:

$$y_i = f(x_i, \Theta) + \varepsilon_i$$

where $f(x_i, \Theta)$ = a non-linear function of Θ,

x = a k-dimensional vector of independent variables,

Θ = a p-dimensional vector of parameters,

ε_i = the error term of the i-th observation,

n = the number of observations.

If we use first order Taylor expansion on the model $f(x_i , \Theta)$ in the neighbourhood of the startvector Θ_0 we get:

$$f(x_i , \Theta) \simeq f(x_i , \Theta_0) + \sum_{j=1}^{p} (\Theta_j - \Theta_{0j}) \left[\frac{\partial f(x_i , \Theta)}{\partial \Theta_j} \right]_{\Theta=\Theta_0}$$

where Θ_{0j} are the components of Θ_0.
This can be written more compactly,

$$f(x , \Theta) \simeq f(x , \Theta_0) + X (\Theta - \Theta_0)$$

$$= f(x , \Theta_0) + X \delta \qquad \ldots \qquad (4.1)$$

where Θ_0 = a vector of first guesses,

$$f(x , \Theta) = \begin{bmatrix} f(x_1 , \Theta) \\ \vdots \\ f(x_i , \Theta) \\ \vdots \\ f(x_n , \Theta) \end{bmatrix}$$

and analogous for $f(x , \Theta_0)$.
Further X is a (nxp) matrix of the form

$$X = \left[\frac{\partial f(x_i , \Theta)}{\partial \Theta_j} \right]_{\Theta=\Theta_0} . \qquad i=1,\ldots,n \; ; \; j=1,\ldots,p$$

The value of δ which minimizes

$$\dot{Q}(x,\Theta) = \sum_{i=1}^{n} \left\{ y_i - f(x_i,\Theta_0) - \sum_{j=1}^{p} (\Theta_j - \Theta_{0j}) \left[\frac{\partial f(x_i,\Theta)}{\partial \Theta_j} \right]_{\Theta=\Theta_0} \right\}^2$$

$$= (y - f(x,\Theta) - X\delta)' (y - f(x,\Theta) - X\delta)$$

is given by

$$\hat{\delta} = (X'X)^{-1} X' e \qquad (4.2)$$

where e is the <u>current</u> vector of the residuals with

$$e_i = \left\{ [y_i - f(x_i,\Theta)] \right\}_{\Theta=\Theta_{(k)}} \quad (k = \text{iteration index})$$

When the matrix X'X is singular, inversion of this matrix is not possible and formula (4.2) is not applicable.
Although we assume indepence of the elements

$$\frac{\partial f(x_i,\Theta)}{\partial \Theta_j}, \qquad i=1,\ldots,n$$

it can accur that, because of computer rounding errors, X'X is almost singular. To avoid this difficulty we use the Newton-Raphson vector

$$\delta = (X'X + C)^{-1} X' e \qquad (4.3)$$

where C is an arbitrary matrix which must prevent singularity of X'X.
Marquardt uses a special form of the C matrix viz.

$$C = \lambda I \tag{4.4}$$

where λ is a parameter that changes during the iteration process.
From (4.3) and (4.4) it is clear that if $\lambda \to 0$ the correction vector δ is the same as in the Gauss-Newton procedure (see 4.2).
A device for improving the numerical aspects of the computing procedure the matrix X'X $\underline{\underline{\text{def}}}$ A and the vector X'e $\underline{\underline{\text{def}}}$ g are scaled.
The matrix A is transformed as follows:

$$A^{*} = D^{-\frac{1}{2}} A D^{\frac{1}{2}}$$

where

$$D^{\frac{1}{2}} = \left\{ \delta_{j_1 j_2} \sqrt{\sum_{i=1}^{n} \frac{\partial f(x_i,\Theta)}{\partial \Theta_{j_1}} \times \frac{\partial f(x_i,\Theta)}{\partial \Theta_{j_2}}} \right\}$$

and $\delta_{j_1 j_2}$ is the Kronecker delta.

By matrix multiplication it can be verified that the elements of the matrix A^{*} are as follows

$$a^{*}_{ii} = 1$$

$$a^{*}_{\substack{jk \\ j \neq k}} = \frac{\sum_{i=1}^{n} \frac{\{\partial f(x_i,\Theta)\}^2}{\partial \Theta_j \, \partial \Theta_k}}{\sqrt{\sum_{i=1}^{n} \frac{\{\partial f(x_i,\Theta)\}^2}{(\partial \Theta_j)^2} \times \sum_{i=1}^{n} \frac{\{\partial f(x_i,\Theta)\}^2}{(\partial \Theta_k)^2}}}$$

When the algorithm of Marquardt converges (even a local minimum may be reached) then

$$\mathcal{E}\left[\left.\frac{\partial f(x_i,\Theta)}{\partial \Theta_j}\right|_{\Theta=\tilde{\Theta}}\right] = 0$$

where $\tilde{\Theta}$ is the vector of parameters of the minimum reached, and then it can be seen that A^* is the matrix of simple correlation coefficients.

Looking at equation (4.2) we see that $X'\,e = g$ is also to be scaled to the vector g^* which can be done as follows:

$$g^* = D^{-\frac{1}{2}}\, g \,. \qquad (4.5)$$

As can be seen g^* is not a normalised vector. It is well known [3a] that the properties of the gradient methods are not scale invariant. It becomes necessary then to transform in every iteration $\hat{\delta}^*$ to $\hat{\delta}$. The above choice of transformation equation (4.5) leads to an easy back calculation of $\hat{\delta}$ and is also numerical sufficient in preventing overflow in the computer.

After transformation, using (4.3) and (4.4), we have to solve

$$(A^* + \lambda\, I)\ \hat{\delta}^* = g^*$$

where $\hat{\delta}$ defined in (4.2) is

$$\hat{\delta} = D^{-\frac{1}{2}}\ \hat{\delta}^* = D^{-\frac{1}{2}}\ D^{\frac{1}{2}}\ (X'X)^{-1}\ D^{\frac{1}{2}}\ D^{-\frac{1}{2}}\ X'\ e$$

$$= (X'X)^{-1}\ X'\ e \qquad \text{(see 4.2)}$$

If $\lambda \to \infty$ (while X'X is transformed into the matrix of simple correlation coefficients among the $\frac{\partial f(x_i,\Theta)}{\partial \Theta_j}$ so that the diagonal elements are equal to one)

then we have approximately

$$\delta \approx \frac{1}{\lambda} X' e = - \frac{1}{2\lambda} \frac{d Q(x,\Theta)}{d \Theta} \quad \underline{\underline{def}} \quad \frac{1}{2\lambda} \delta_g$$

where δ_g is the negative gradient (vector of steepest descent) of $Q(x,\Theta)$.

The theoretical basis of this algorithm is formed by the following three theorems.

Theorem 4.1

Let $\lambda \geq 0$ be arbitrary and let δ_0 satisfy the equation

$$(X'X + \lambda I)\, \delta = X' e \quad \underline{\underline{def}} \quad g. \qquad (4.6)$$

Then δ_0 minimizes $\dot{Q}(x,\Theta)$ everywhere except within the ellipsoid Ω consisting of all points Θ such that $\dot{Q}(x,\Theta) = \dot{Q}(x,\Theta_0 + \delta_0)$. In particular δ_0 determines the unique minimum of $\dot{Q}(x,\Theta)$ on and within the sphere Φ, centered at Θ_0, whose radius $\|\delta\|$ satisfies

$$\|\delta\|^2 = \|\delta_0\|^2 , \quad \text{where} \quad \|\delta\|^2 = \sum_{j=1}^{p} \delta_j^2 .$$

This can be clarified as follows (see also figure 4.1). From linear least squares theory it is known that contours of constant $\dot{Q}(x,\Theta)$ on the parameter locus are ellipsoids because of the assumption of normality.

Ellipsoids in the parameter space generate constant levels of $Q(x,\Theta)$ such that Θ_1, Θ_2 and Θ_3 are points that are outside, on, and within a given ellipsoid, respectively, then

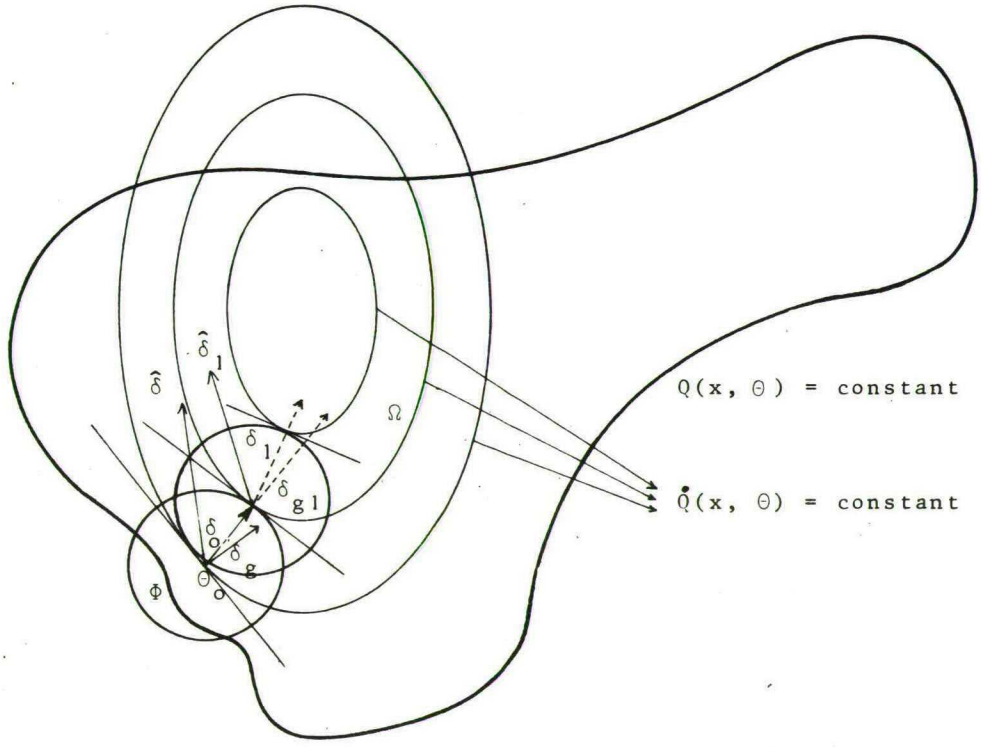


figure 4.1

Diagram showing the sphere Φ and the ellipsoid Ω of theorem 4.1

$$\dot{Q}(x,\Theta_1) > \dot{Q}(x,\Theta_2) > \dot{Q}(x,\Theta_3).$$

All points Θ on Ω, in particular the point $\Theta_0 + \delta_0$, minimize $\dot{Q}(x,\Theta)$ outside or on Ω. We can show that the sphere Φ is tangent to the ellipsoid Ω.

$\dot{Q}(x,\Theta)$ can be written as

$$(e - X\,\delta)'\,(e - X\,\delta).$$

The gradient of $\dot{Q}(x,\Theta)$ is given by

$$\frac{\partial}{\partial\,\delta}\,(e - X\,\delta)'(e - x\,\delta) = -\,2\,X'\,(e - X\,\delta) \qquad (4.7)$$

so that when evaluated at $\Theta = \Theta_0 + \delta_0$ the gradient is

$$-\,2\,(X'\,e - X'X\,\delta_0).$$

A rearrangement of (4.6) gives us

$$X'X\,\delta_0 = X'\,e - \lambda\,I\,\delta_0$$

which when substituted in (4.7) gives the result that the gradient of $\dot{Q}(x,\Theta)$ at the point $\Theta_0 + \delta_0$ is given by $-\,2\,\lambda\,\delta_0$. Since $\lambda \geq 0$ this means that the gradient of $\dot{Q}(x,\Theta)$ at $\Theta_0 + \delta_0$ is collinear with the vector δ_0, but having opposite direction. So the steepest <u>ascent</u> vector of $\dot{Q}(x,\Theta)$ is perpendicular to the plane tangent to the ellipsoid Ω at the point $\Theta_0 + \delta_0$ and pointing <u>out</u> from Ω.

Since δ_0 is the vector originating at the center of the sphere Φ and terminating at the ellipsoid Ω at the point

$\Theta_0 + \delta_0$, it follows that Φ and Ω are externally tangent at $\Theta_0 + \delta_0$.
As a conclusion of this geometrical proof we can say that δ_0 determines the unique minimum of $Q(x,\Theta)$ on and within Φ because Φ and Ω are tangent at only one point.
We now choose λ so that $Q(x,\Theta)$ is a good approximation for $Q(x,\Theta)$ within a neighbourhood around Θ_0:
We must keep in mind that δ_0 minimizes $Q(x,\Theta)$ on and within Φ rather than just on Φ.

Theorem 4.2

Let $\delta(\lambda)$ be the solution of (4.6) for a given λ. Then $\|\delta(\lambda)\|^2$ is a continuous decreasing function of λ such that as $\lambda \to \infty$, $\|\delta(\lambda)\|^2 \to 0$.

Theorem 4.3

Let α be the angle between δ_0 and δ_g. Then α is a continuous monotone decreasing function of λ such that as $\lambda \to \infty$, $\alpha \to 0$.
Further, since δ_g is independent of λ then δ_0 rotates toward δ_g as $\lambda \to \infty$.
For proofs of this theorems we refer to Marquardt [11].
For theorem 2 and 3 the two dimensional case will be illustrated in figure 4.2.

The algorithm of the procedure of Marquardt runs as follows:

(0) $\Theta^{(o)}$ is a given starting point
λ is a given value
k is iteration number = 1

(1) calculate $X^{(k)}$

(2) form $(X'X)^{(k)}$

(3) form $g^{(k)} = (X'e)^{(k)}$

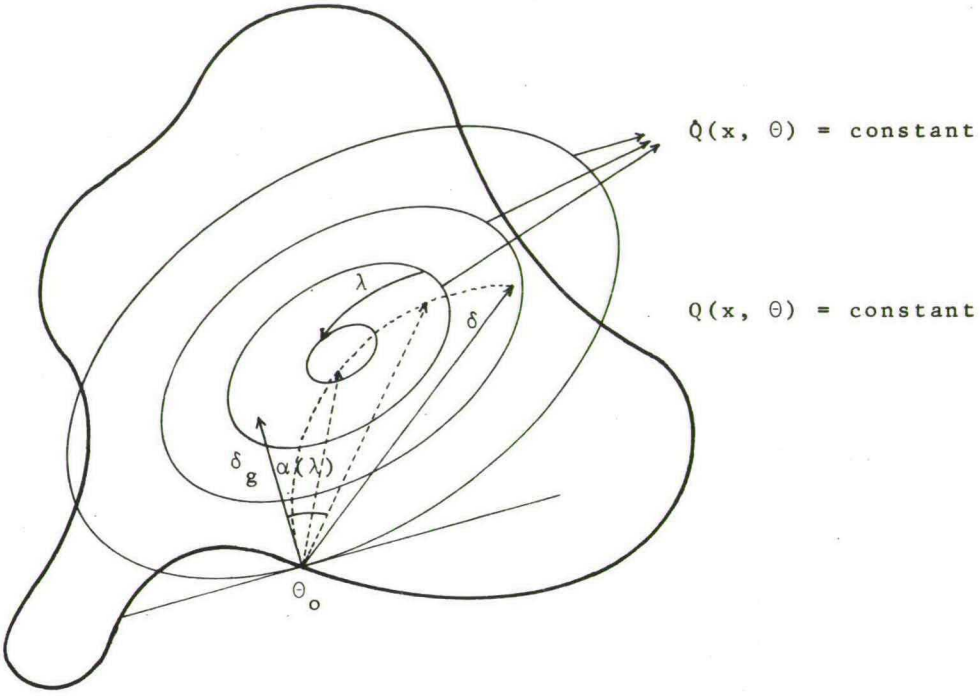


figure 4.2.

Diagram showing the decreasing quadratic norm of the vector $\delta(\lambda)$ and the decreasing angle α for increasing λ.

(4) calculate $\delta^{(k)}$ from $[X'X^{(k)} + \lambda I]\ \delta^{(k)} = g^{(k)}$

(5) calculate $Q[x, \Theta^{(k)} + \delta^{(k)}]$

(6) if $Q[x, \Theta^{(k)} + \delta^{(k)}] < Q_{min}^{(k-1)}$

(where $Q_{min}^{(k-1)}$ denotes the minimum function value in the preceding iteration) then

$\Theta^{(k+1)} = \Theta^{(k)} + \delta^{(k)}$ and $\lambda = \frac{\lambda}{\nu}$.

Let $\nu > 0$ be, say 10.

Set k = k+1 and go to (1)

(7) If $Q\,[x,\Theta^{(k)} + \delta^{(k)}] \geq Q_{min}^{(k-1)}$ then

$\lambda = \lambda * \nu$ and go to (4)

Convergence of this algorithm will be reached when for all j

$$\max_j \left| \frac{\Theta_j^{(k+1)} - \Theta_j^{(k)}}{\Theta_j^{(k)} + 0.00001} \right| \leq E\ P\ S$$

where $\Theta_j^{(k)}$ is the j-th element of the parameter vector Θ in the k-th iteration.

and E P S = 0.00001

Remark: In the neighbourhood of the minimum the approximation of $Q\ (x,\Theta)$ by a quadratic form will become better and so λ can be decreased more and more. The result is that the Marquardt procedure becomes more alike the Gauss-Newton procedure of which it is known that is has a good convergence in the neighbourhood of the minimum.

5. NUMERICAL RESULTS.

5.1 Introduction.

The methods described were tested for various initial estimates on two functions.
They are the Weibull function defined by

$$f(\Theta_1,\Theta_2,\Theta_3) = \sum_{i=1}^{99} \left(e^{-\frac{1}{\Theta_1}(x_{1i}-\Theta_3)^{\Theta_2}} - x_{2i}\right)^2$$

where the x_{2i} and x_{1i} are perfect data generated for the 99 points corresponding to x_{2i} = .01 to .99 in steps of 0.01 and $x_{1i} = (-50 * \ln x_{2i})^{2/3} + 25$.
The minimum of this function is zero for the values Θ_1=50, Θ_2=1.5, Θ_3=25.
The initial estimates are those suggested by Shanno [16].

The Rosenbrock function defined by

$$f(\Theta_1,\Theta_2)=100\ (\Theta_2-\Theta_1^2)^2 + (1-\Theta_1)^2\ ,$$

was also used, with the initial estimates suggested by Leon [9].
The minimum of this function is again zero, for the values Θ_1=1 and Θ_2=1.
In the tables, the number of iterations is the number of times $H^{(k)}$ is updated, and the number of evaluations the actual number of function evaluations used.
For the quasi Newton methods convergence was said to be reached when for all i

$$\left|\frac{\theta_i^{(k+1)} - \theta_i^{(k)}}{\theta_i^{(k)} + 0.001}\right| \leq EPS \quad \text{and} \quad \frac{|g_i^{(k)}|}{|\theta_i^{(k)} + 0.001|} \leq EPS$$

The following versions for selecting t discussed in section 2.1.4 were tested:

JSCALE = 0 : $t = \infty$

JSCALE = 1 : $t = \frac{2\,\alpha_{(k)}^{-1}}{\alpha_{(k)}}$

JSCALE = 2 : t=1 (Fletcher-Powell)

JSCALE = 3 : t=0 (Barnes-Rosen)

JSCALE = 4 : constant norm

For the Marquardt method the convergence criterion was reached when for all i

$$\left|\frac{\theta_i^{(k+1)} - \theta_i^{(k)}}{\theta_i^{(k)} + 0.00001}\right| \leq EPS.$$

It is evident that the two convergence criteria in the quasi-Newton methods are stronger than the one of Marquardt's procedure. Therefore in the numerical examples of Marquardt's method

EPS = 0.00001

and in the quasi-Newton methods

for the Rosenbrock function EPS = 0.001 or 0.0001
for the Weibull function EPS = 0.0001 or 0.00001

The smaller values of EPS are taken because in the computer-program the procedure will be finished when the steplength becomes smaller than EPS.

We also studied the four parameter logistic function defined by

$$f(t;\theta_1, \theta_2, \theta_3, \theta_4) = \theta_1 + \frac{\theta_2}{1 + \theta_3 e^{-\theta_4 t}} \qquad (5.1)$$

and we minimized the following function

$$g(t;\theta_1, \theta_2, \theta_3, \theta_4) = \sum_{t=1}^{N} \left\{ y_t - f(t;\theta_1, \theta_2, \theta_3, \theta_4) \right\}^2 \qquad (5.2)$$

where N = sample size,

y_t = the values of the dependent variables,

t = time variable.

This function arises in many economic contexts where growth phenomena are studied.
Further details will be given in section 5.4.

5.2

The numerical results for the quasi-Newton methods of the Rosenbrock and Weibull function are given in Table 1 (Rosenbrock function with cubic interpolation), Table 2 (Rosenbrock function with both quadratic interpolation procedures), Table 3 (Weibull function with cubic interpolation), and Table 4 (Weibull function with quadratic interpolation).

5.3

The numerical results for the Marquardt method of the

Rosenbrock and Weibull functions are given in Table 5.
Of the two versions given is M 1 the original Marquardt procedure and in the version M 2 a technical change to decrease the number of function evaluations, is incorporated in the computer program.
For details we refer to an internal report at the Tilburg School for Economics written by Dieben L. [4].
The results of the Marquardt procedure in table 5 are documented by Dieben L.

Table 1

Quasi-Newton procedure -Rosenbrock function- cubic interpolation

Initial values	JSCALE				
	0	1	2	3	4
EPS=0.0001					
4 ; 16	28*,102**	28,107	32,129	28,93	28,107
3 ;- 3	61,195	25,88	68,282	54,184	24,90
0 ; 3	19,60	18,63	19,63	18,57	17,55
-3 ; 9	26,89	23,94	24,98	19,63	27,100
EPS=0.001					
8 ; 6	18,68	20,74	20,77	22,73	23,81
3 ; 3	17,63	16,61	16,66	14,50	15,57
3 ; 1	8,32	7,30	10,52	8,32	8,34
3 ; 0	18,63	20,75	17,72	19,67	19,66
2 ;- 1	33,122	35,138	41,184	37,126	33,122
2 ;- 2	34,125	39,145	41,165	37,124	33,128
1.489 ;-2.547	37,124	41,144	42,177	36,122	36,134
1 ; 3	15,56	15,55	17,66	14,52	16,57
1 ;-1.2	54,183	28,100	57,225	54,180	26,105
0.639;-0.221	27,92	27,96	31.117	27,93	21,76
0 ; 2	19,60	17,60	18,65	17,54	13,42
0 ;- 3	19,59	17,59	17,60	17,54	17,56
-1.2 ; 1	13,64	15,52	12,49	13,50	15,52
- 2 ; 16	35,129	37,148	38,163	40,147	35,133
- 2 ; 2	15,54	16,57	22,82	21,66	18,66
- 2 ;- 2	44,156	44,168	53,244	53,182	44,169
- 3 ; 5	20,82	20,75	21,92	19,76	19,68
- 3 ; 0	15,60	18,69	21,79	13,51	15,67
-3.635;5.621	29,98	29,110	31,134	28,103	27,98

* : Number of iterations

** : Number of function evaluations

Quasi-Newton procedure - Rosenbrock function - Quadratic interpolation

Table 2

JSCALE / Initial values	with equidistant points					with non-equidistant points				
	0	1	2	3	4	0	1	2	3	4
EPS = 0.0001										
4;16	25*;306**	27,349	29,400	27,401	29,328	26,183	26,211	26,252	29,218	28,216
3;-3	15,245	16,241	16,232	16,228	16,262	14,138	15,134	13,144	15,134	15,133
0; 3	15,230	failed***	18,234	15,275	15,227	16,121	16,115	15,143	15,126	13,100
-3 ; 9	32,398	28,436	34,499	36,452	33,480	31,259	31,263	33,314	32,292	31,256
EPS = 0.0001										
8; 6	11,181	9,146	18,231	14,151	8,137	20,180	22,234	23,210	19,200	19,182
3; 3	17,207	17,351	16,238	17,225	18,193	17,133	16,147	16,155	17,135	15,119
3; 1	7,91	7,89	10,121	7,94	7,132	7,99	6,77	8,107	7,94	7,89
3; 0	28,344	24,312	31,403	30,399	25,373	12,141	12,164	12,126	12,140	12,136
2;-1	18,278	18,240	17,262	18,316	18,243	18,178	19,171	16,163	16,192	19,190
2;-2	25,293	27,224	22,334	25,338	23,297	14,135	17,167	16,156	15,151	14,137
1.489;-2.547	15,225	17,177	17,203	18,187	17,181	16,158	16,177	16,154	14,152	16,186
1; 3	12,193	11,175	14,196	15,194	11,171	11,119	13,142	14,148	13,114	12,108
1;-1.2	17,216	17,229	19,263	14,249	17,269	16,148	23,188	16,151	19,168	17,123
0.639;-0.221	21,231	19,261	15,193	21,338	17,198	5,70	5,58	13,104	5,73	16,124
0; 2	failed	15,213	16,232	failed	15,249	15,111	13,90	15,137	13,94	14,111
0;-3	15,219	15,236	17,216	15,157	15,288	15,131	13,93	13,109	13,104	13,93
-1.2; 1	20,308	24,337	20,415	22,332	21,328	20,165	20,161	19,178	20,169	21,171
-2 ;16	14,452	13,451	22,551	37,589	36,494	35,307	38,326	43,399	36,349	37,328
-2 ; 2	25,293	27,324	22,334	25,338	23,297	26,213	23,205	24,242	26,229	27,222
-2 ;-2	18,214	21,267	19,276	21,292	21,260	19,156	19,166	19,182	18,168	19,165
-3 ; 5	28,291	27,486	32,451	27,441	31,374	29,222	29,247	26,238	29,256	33,251
-3 ; 0	15,236	15,186	20,250	15,210	16,217	20,188	22,220	20,202	22,213	22,194
-3.635; 5.621	32,356	30,406	32,322	31,417	31,388	29,228	28,230	34,299	32,273	31,237

* : number of iterations

** : number of function evaluations

***: point of convergence not reached.

Table 3

Quasi-Newton procedure -Weibull function- cubic interpolation

Initial values	EPS	JSCALE				
		0	1	2	3	4
250,0.5,5	0.0001	38*,163**	38,205	(8,35)***	39,136	(13,74)
	0.0001	39,165	38,205	87,489	40,143	42,201
100,3,12.5	0.0001	28,128	31,165	40,214	41,176	(14,101)
	0.00001	28,128	31,165	40,214	41,176	28,146
75,0.5,10	0.0001	24,99	25,116	30,138	23,81	22,92
	0.00001	25,101	26,120	31,142	24,85	23,95
25,0.5,10	0.0001	21,101	21,100	26,117	20,107	19,95
	0.00001	22,105	22,109	26,117	21,111	20,97
5,0.15,2.5	0.0001	18,88	failed****	25,102	18,80	19,93
	0.00001	19,90	23,100	26,105	18,80	19,93

* = number of iterations

** = number of functie evaluations

*** = convergence to stationary point, not minimum

**** = point of convergence not reached

Table 4

Quasi-Newton procedure -Weibull function- quadratic interpolation

Version E = with equidistant points
Version NE = with points not necessarily equidistant

Initial values	Version	EPS	ISCALE 0	1	2	3	4 .
250;0.5; 5	NE	0.0001	32,261[1),2)]	(8,64)[3)]	(7,61)	(8,68)	(9,67)
	NE	0.00001	.,.	35,255	110,1048	40,254	38,277
100;3;12.5	E	0.0001	24,335	$\alpha < 0$[4)]	max[5)]	29,451	28,400
	NE	0.0001	23,208	26,210	43,389	27,215	(7,99)
	NE	0.00001	.,.	.,.	.,.	.,.	25,221
75;0.5;10	E	0.0001	(7,217)	11,216	49,445	failed[6)]	(7,217)
	E	0.00001	23,384	20,267	50,451	24,298	$\alpha < 0$
	NE	0.0001	19,142	19,174	33,261	23,183	23,160
25;0.5;10	E	0.0001	22,257	22,244	37,364	23,254	23,238
	NE	0.0001	21,160	20,129	29,236	23,155	23,183
5;0.5;2.5	E	0.0001	18,233	25,330	31,425	19,241	24,930
	NE	0.0001	27,211	$\alpha < 0$	19,177	27,201	18,139

1) : number of iterations
2) : number of function evaluations
3) : convergence to stationary point, not minimum
4) : steplength negative
5) : no convergence after 50 iterations
6) : point of convergence not reached

Table 5

Marquardt procedure -Rosenbrock and Weybull function-

Rosenbrock function	Version			Version	
Initial value	M_1	M_2	Initial value	M_1	M_2
4 ; 16	17,52[1)2)]	17,47	1 ;-1.2	5,16	5,11
3 ;- 3	8,25	8,20	0.639;-0.221	9,28	9,23
0 ; 3	5,16	5,11	0 ; 2	5,16	5,11
-3 ; 9	45,139	45,131	0 ;-3	5,16	5,11
8 ; 6	28,85	28,80	-1.2 ; 1	23,72	23,65
3 ; 3	11,34	11,29	- 2 ; 16	37,116	37,107
3 ; 1	10,31	10,26	- 2 ; 2	26,81	26,74
3 ; 0	9,28	9,23	- 2 ;-2	23,73	23,65
2 ;- 1	5,16	5,11	- 3 ; 5	36,112	36,104
2 ;- 2	5,16	5,11	- 3 ; 0	25,80	25,72
1.489;-2.547	5,16	5,11	-3.635 ; 5.621	37,115	37,107
1 ; 3	6,19	6,14			

Weybull function	Version			Version	
Initial value	M_1	M_2	Initial value	M_1	M_2
100; 3;12.5	3)	3)	25;0.5,10	15,44	15,37
75;0.5;10	20,61	20,53	5;0.15,2.5	15,50	15,40

1) : number of iterations

2) : number of function evaluations

3) : convergence to stationary point, not minimum

5.4

As already given before the formula for the logistic curve as a function of time is

$$f(t;\Theta_1,\Theta_2,\Theta_3,\Theta_4) = \Theta_1 + \frac{\Theta_2}{1+\Theta_3\, e^{-\Theta_4 t}} \qquad (5.3)$$

In the error model formulation the observed values $y_1,\ldots,y_N$ are assumed to differ from the trend by uncorrelated random errors, and we may establish a least squares criterion for estimating the parameters $\Theta_1,\Theta_2,\Theta_3$ and Θ_4.
Thus we require that the expression

$$g(t;\Theta_1,\Theta_2,\Theta_3,\Theta_4) = \sum_{t=1}^{N} \left(y_t - \Theta_1 - \frac{\Theta_2}{1+\Theta_3\, e^{-\Theta_4 t}}\right)^2 \qquad (5.4)$$

be minimized with respect to $\Theta_1,\Theta_2,\Theta_3$, and Θ_4.

The minimizing values $\hat{\Theta}_1,\hat{\Theta}_2,\hat{\Theta}_3$, and $\hat{\Theta}_4$ will be the least squares estimates of the parameters $\Theta_1,\Theta_2,\Theta_3$, and Θ_4 respectivily. This computational problem may often be solved iteratively.

5.4.1

Using a Taylor series expansion of $f(t;\Theta)$ about suitable values Θ^o and considering the first terms only:

$$f(t;\Theta) = f(t;\Theta^o) + \sum_{j=1}^{4} (\Theta_j - \Theta_j^o) \left.\frac{\partial f(t;\Theta)}{\partial \Theta_j}\right]_{\Theta^o} + R \qquad (5.5)$$

where, Θ is the vector with elements Θ_1, Θ_2, Θ_3, Θ_4

Θ^o is the vector with elements Θ_1^o, Θ_2^o, Θ_3^o, Θ_4^o

$f(t;\Theta) = f(t;\Theta_1, \Theta_2, \Theta_3, \Theta_4)$

$f(t;\Theta^o) = f(t;\Theta_1^o, \Theta_2^o, \Theta_3^o, \Theta_4^o)$

R is a remainder term.

Given initial estimates of Θ_1^o, Θ_2^o, Θ_3^o, and Θ_4^o a linear least squares technique can be applied to obtain estimates Θ_1^1, Θ_2^1, Θ_3^1, Θ_4^1.

Substituting (5.5) without remainder term in (5.4) the resulting linear equations for $\Theta_i - \Theta_i^o$ are

$$\sum_{t=1}^{N} \frac{\partial f(t;\Theta^o)}{\partial \Theta^o} \left[\frac{\partial f(t;\Theta^o)}{\partial \Theta_o}\right]' (\Theta - \Theta^o) =$$

$$\sum_{t=1}^{N} [y_t - f(t;\Theta^o)] \frac{\partial f(t;\Theta^o)}{\partial \Theta^o} \qquad (5.6)$$

where

$\Theta - \Theta^o$ is a column vector with elements $\Theta_i - \Theta_i^o$

$\frac{\partial f(t;\Theta^o)}{\partial \Theta^o}$ is a column vector whose elements are the partial derivatives with respect to Θ_1^o, Θ_2^o, Θ_3^o, and Θ_4^o evaluated at the initial points.

The procedure can be repeated using Θ_1^1, Θ_2^1, Θ_3^1, and Θ_4^1 in place of Θ_1^o, Θ_2^o, Θ_3^o, Θ_4^o to obtain more exact solutions, and so on.

The sequence so obtained converges in many cases.

5.4.2

Another way of developing an iterative procedure is to take the partial derivatives of (5.4) with respect to Θ_1, Θ_2, Θ_3, and Θ_4 and set them equal to zero, yielding.

$$\sum_{t=1}^{N} f(t;\Theta) \frac{\partial f(t;\Theta)}{\partial \Theta_i} = \sum_{t=1}^{N} y_t \frac{\partial f(t;\Theta)}{\partial \Theta_i} \quad i=1,2,3,4 \qquad (5.7)$$

The Taylor series expansion of the derivative vector is

$$\frac{\partial f(t;\Theta)}{\partial \Theta} = \frac{\partial f(t;\Theta^o)}{\partial \Theta^o} + \sum_{j=1}^{N} (\Theta_j - \Theta_j^o) \frac{\partial^2 f(t;\Theta)}{\partial \Theta_i \partial \Theta_j} \Bigg]_{\Theta_1^o, \Theta_2^o, \Theta_3^o, \Theta_4^o} + R$$

or written elsewhise

$$\frac{\partial f(t;\Theta)}{\partial \Theta} = \frac{\partial f(t;\Theta^o)}{\partial \Theta^o} + (\Theta - \Theta^o) \frac{\partial^2 f(t;\Theta^o)}{\partial \Theta^o (\partial \Theta^o)'} \qquad (5.8)$$

When substituting (5.8) and (5.5) without remainder term in (5.7) we obtain as resulting equations for $(\Theta - \Theta^o)$.

$$\left\{ \sum_{t=1}^{N} \frac{\partial f(t;\Theta^o)}{\partial \Theta^o} \left[\frac{\partial f(t;\Theta^o)}{\partial \Theta^o} \right]' - \sum_{t=1}^{N} [y_t - f(t;\Theta^o)] \frac{\partial^2 f(t;\ ^o)}{\partial \Theta^o (\partial \Theta^o)'} \right\}$$

$$(\Theta - \Theta^o) =$$

$$= \sum_{t=1}^{N} [y_t - f(t;\Theta^o)] \frac{\partial f(t;\Theta^o)}{\partial \Theta^o} \qquad (5.9)$$

where

$$\frac{\partial^2 f(t;\Theta^o)}{\partial\Theta^o(\partial\Theta^o)'} = \left[\frac{\partial^2 f(t;\Theta)}{\partial\Theta_i\partial\Theta_j}\right]_{\Theta_1^o,\ \Theta_2^o,\ \Theta_3^o,\ \Theta_4^o} \tag{5.10}$$

If the random terms are normally and independent distributed with constant variance the method of least squares is equivalent with the maximum-likelihood method.
If

$$\left[\frac{\partial^2 f(t;\Theta)}{\partial\Theta_i\partial\Theta_j}\right]_{\hat{\Theta}_1,\ \hat{\Theta}_2,\ \hat{\Theta}_3,\ \hat{\Theta}_4} \tag{5.11}$$

where $\hat{\Theta}_1,\ldots,\hat{\Theta}_4$ are consistent estimates, is replaced by its expected value the procedure is called the method of scoring. It can be proved that the expectation of (5.11) converges in

in probability to $\mathcal{E}\left[\frac{\partial^2 f(t;\Theta)}{\partial\Theta_i\partial\Theta_j}\right]_{\Theta\,=\,\hat{\Theta}}$

where $\hat{\Theta}$ is the vector of true parameters.
The expectation of (5.11) is called the information matrix (see Rao, C.R. [14]).
At the final stage when stable values are reached, the information matrix may be computed at the estimated values for obtaining the variances and covariances of the estimates.

5.4.3
Marquardt's algorithm is a compromise between the linearization of the function f(t;Θ) and the method of steepest descent. Further details can be found in chapter 4.

5.4.4

For the Quasi-Newton methods we refer to chapter 2 and 3.

5.4.5

Numerical results of a four parameter logistic growth function (5.3) can be found in Table 6 below.
The data used are documented by Oliver, F.R. [12].

The values of the calculated parameters for several initial values are only given (see Table 7) for the method of Marquardt. We did not get results by the Quasi-Newton procedure, because, when taking an unit step in the direction of the gradient from a point not near the minimum, the calculation of the function value causes overflow in the computer.

Table 6

Agricultural tractors in Great Britain (thousands)

t	year	number of tractors	t	year	number of tractors
1	1950	296	9	1958	414
2	1951	282	10	1959	432
3	1952	305	11	1960	443
4	1953	327	12	1961	452
5	1954	346	13	1962	453
6	1955	368	14	1963	465
7	1956	379	15	1964	475
8	1957	400	16	1965	470

Table 7

Marquardt procedure - logistic curve

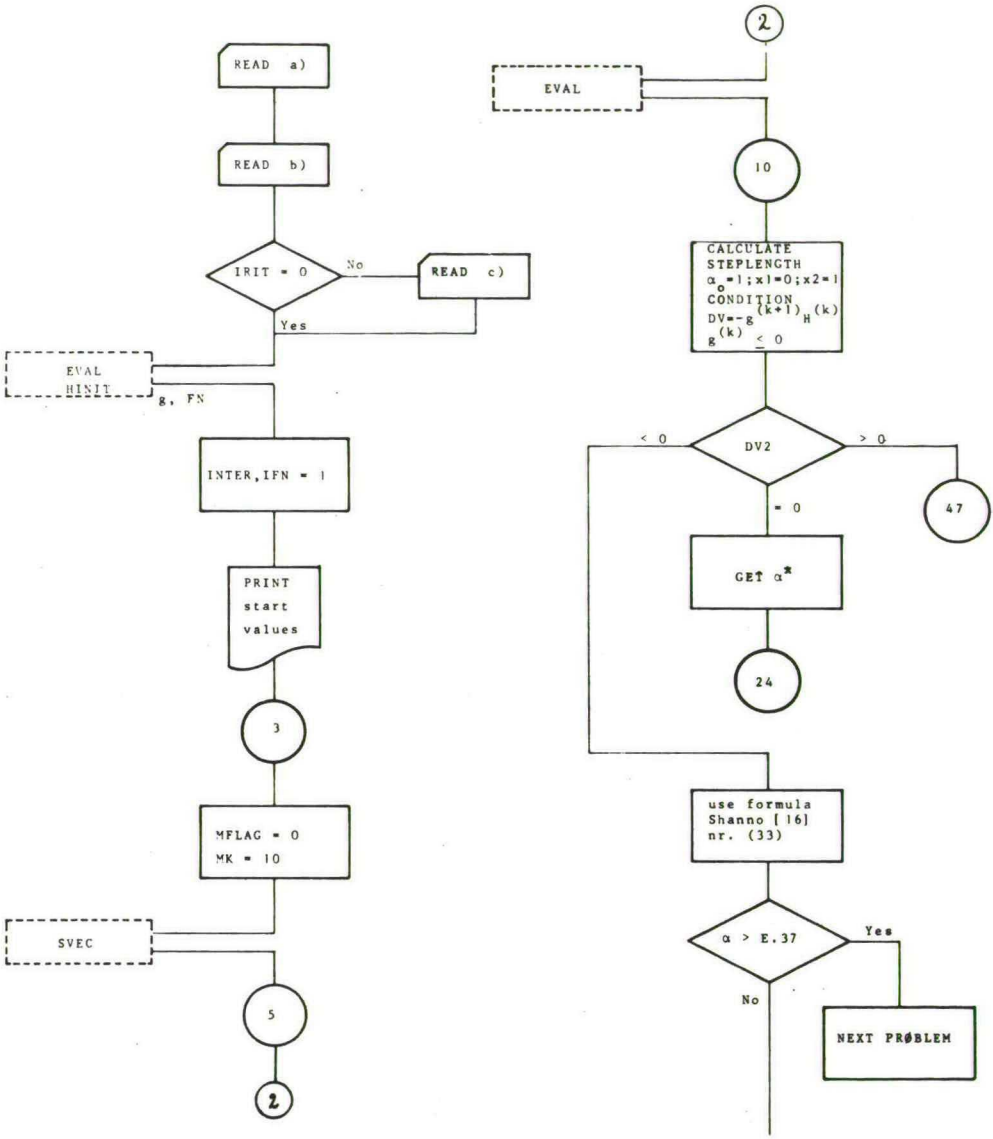
Initial values	calculated parameter values				number of	
	$\hat{\Theta}_1$	$\hat{\Theta}_2$	$\hat{\Theta}_3$	$\hat{\Theta}_4$	iterations	function evaluations
300;200;15;0.85	250.9	228.4	8.119	0.3404	8	19
200;300; 5;0.10	250.9	228.4	8.119	0.3405	10	25
175;325;15;0.85	250.9	228.4	8.119	0.3405	21	58

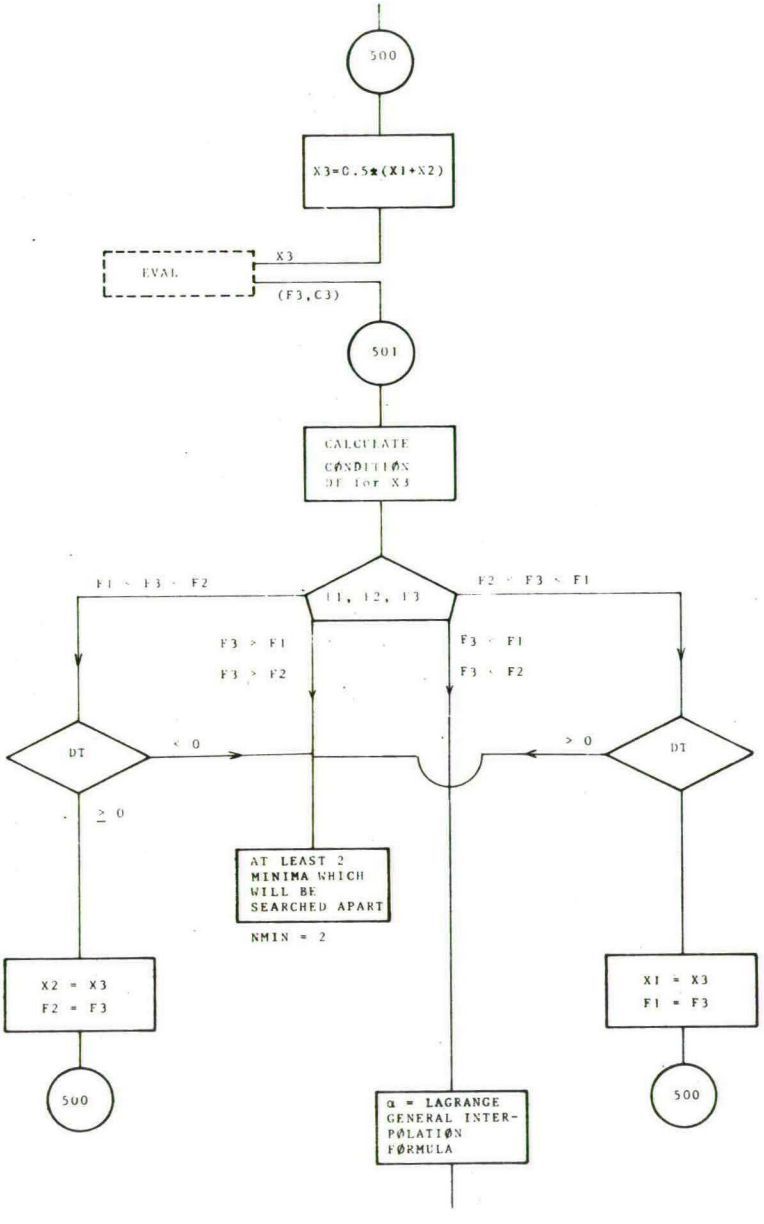
6. Conclusion.

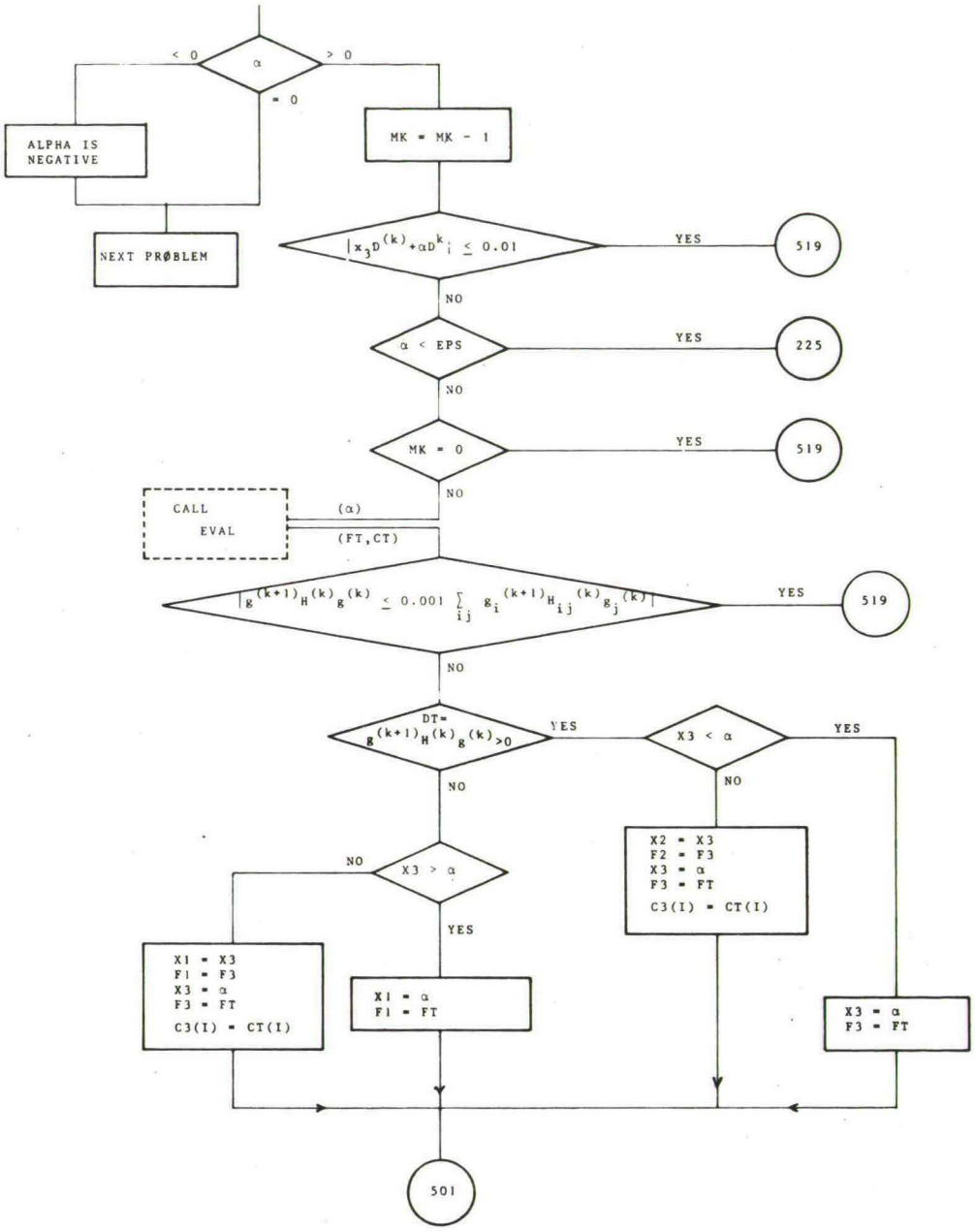
Comparing, for the Quasi-Newton procedure the two quadratic interpolation versions (equidistant (E)-non equidistant (NE) points) we found that, for both the Rosenbrock and Weibull test functions, the version with points not necessarily equidistant, gives the better results in the number of iterations and function evaluations.
The results of the cubic interpolation procedure are, for the two test functions, in most cases better than the results of the quadratic interpolation (version NE).
As can be seen from the tables we may conclude that dependent on the initial estimates sometimes Marquardt's procedure, sometimes the Quasi-Newton procedure with cubic interpolation is better, with a slight preference for Marquardt's method.

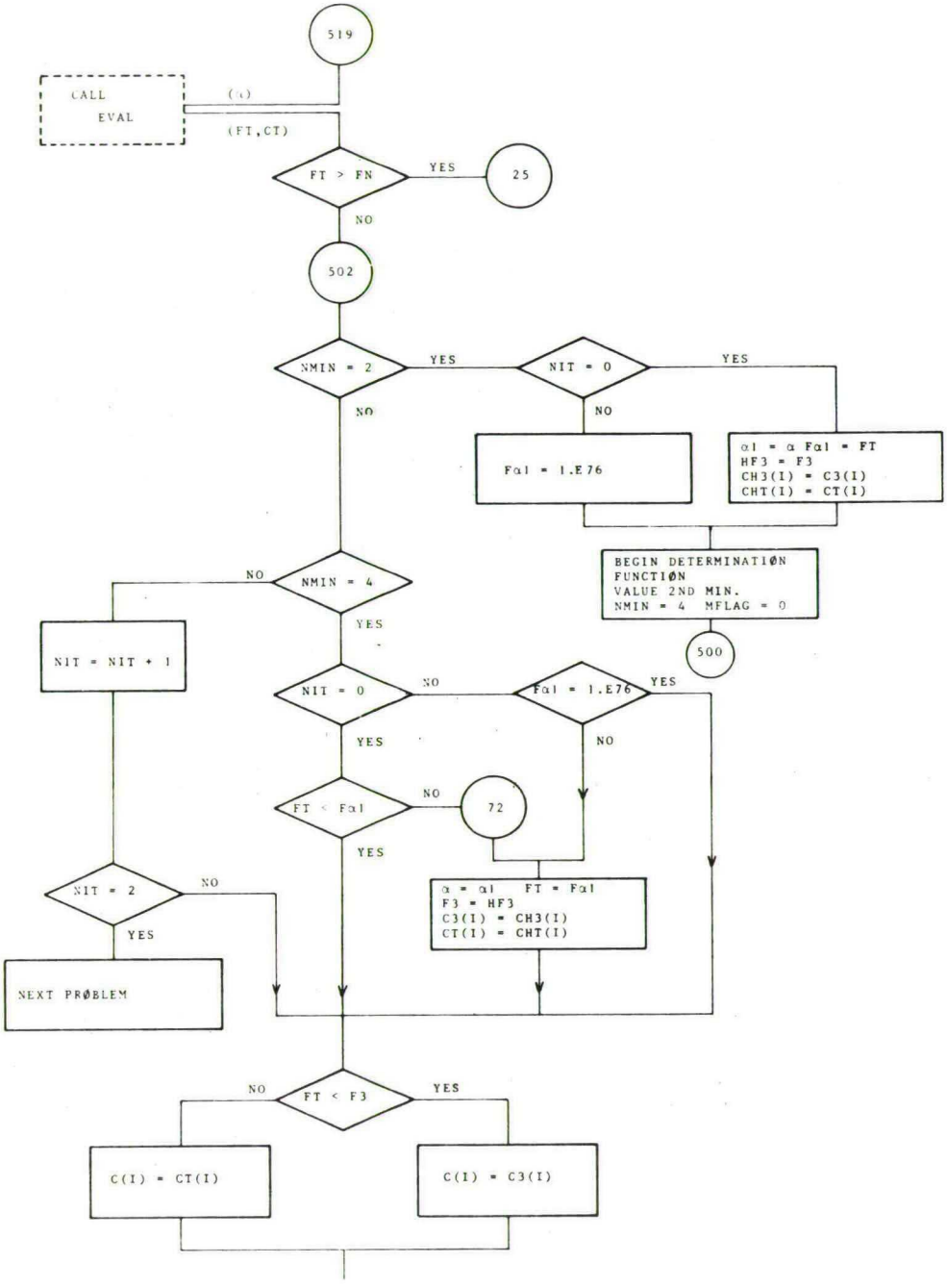
When exponential functions have to be minimized it is preferable to choose the algorithm of Marquardt which uses a scaling procedure, while the Quasi-Newton algorithm often causes overflow in the computer.

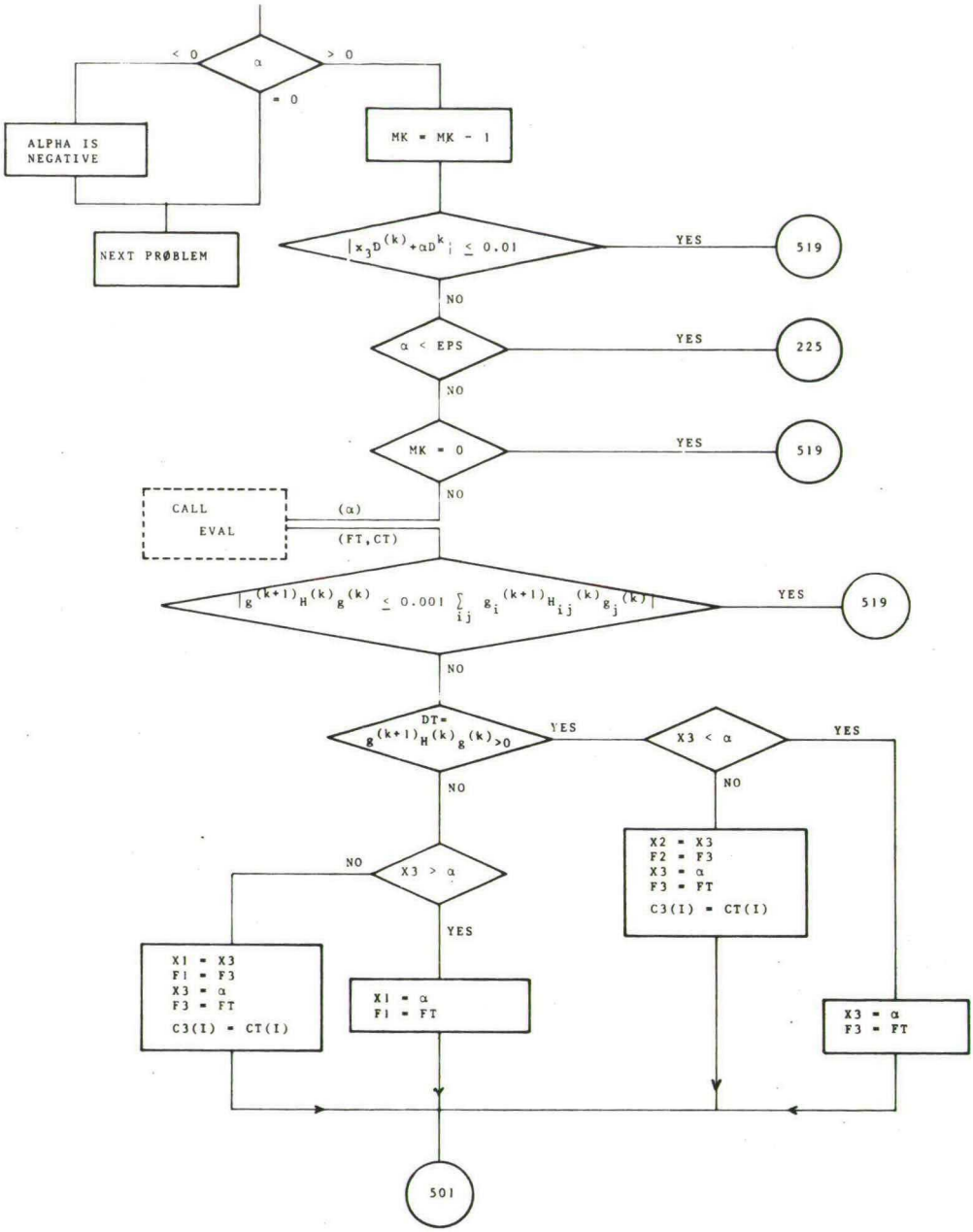
7. Flowchart of the algorithm in chapter three.

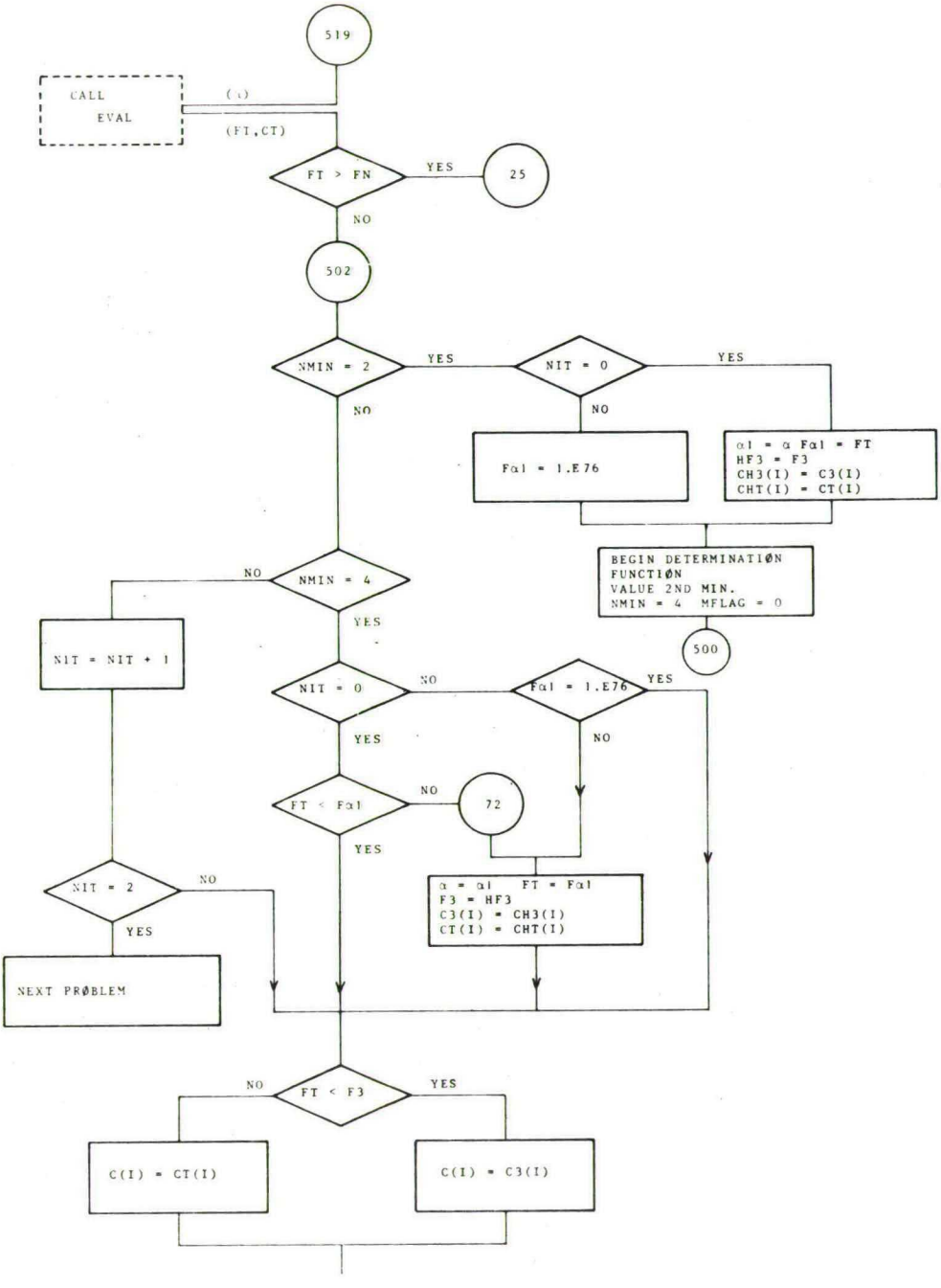


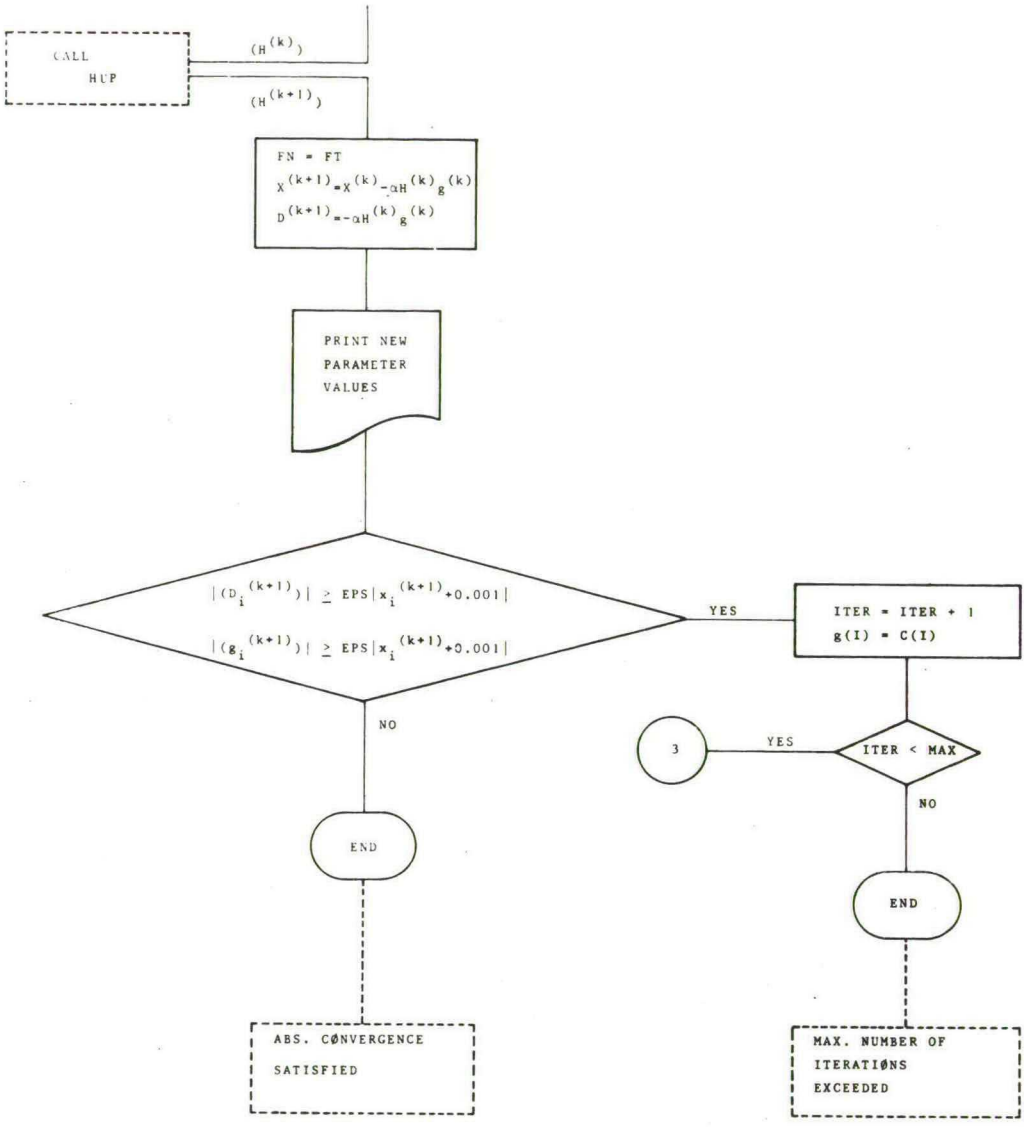

500
X3=0.5*(X1+X2)
EVAL
X3
(F3,C3)
501
CALCULATE
CONDITION
OF for X3
F1 < F3 < F2
I1, I2, I3
F2 < F3 < F1
F3 > F1
F3 > F2
F3 < F1
F3 < F2
DT
< 0
> 0
DT
≥ 0
AT LEAST 2
MINIMA WHICH
WILL BE
SEARCHED APART
NMIN = 2
X2 = X3
F2 = F3
X1 = X3
F1 = F3
500
α = LAGRANGE
GENERAL INTER-
POLATION
FORMULA
500

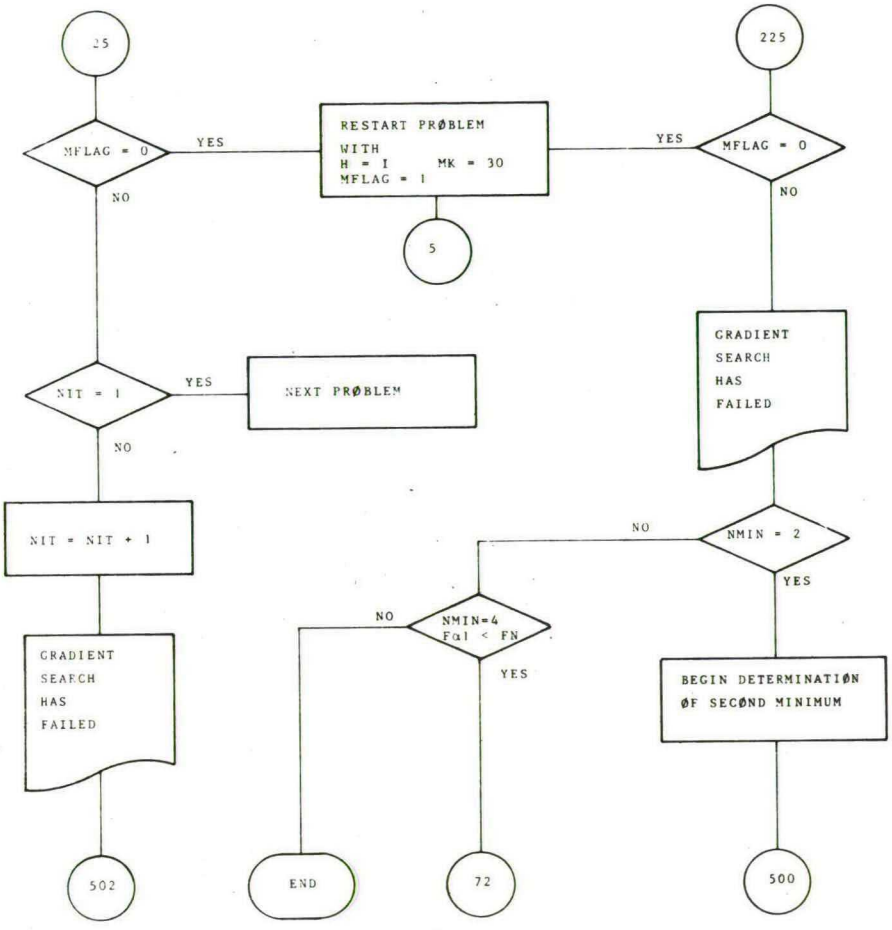

α
< 0
> 0
= 0
ALPHA IS NEGATIVE
NEXT PRØBLEM
MK = MK - 1
$|x_3 D^{(k)} + \alpha D^k{}_i \leq 0.01$
YES
519
NO
α < EPS
YES
225
NO
MK = 0
YES
519
NO
CALL EVAL
(α)
(FT,CT)
$|g^{(k+1)} H^{(k)} g^{(k)} \leq 0.001 \sum_{ij} g_i^{(k+1)} H_{ij}^{(k)} g_j^{(k)}|$
YES
519
NO
$DT = g^{(k+1)} H^{(k)} g^{(k)} > 0$
YES
NO
X3 < α
YES
NO
X2 = X3
F2 = F3
X3 = α
F3 = FT
C3(I) = CT(I)
X3 > α
NO
YES
X1 = X3
F1 = F3
X3 = α
F3 = FT
C3(I) = CT(I)
X1 = α
F1 = FT
X3 = α
F3 = FT
501


519
CALL EVAL
(α)
(FT,CT)
FT > FN
YES
25
NO
502
NMIN = 2
YES
NIT = 0
YES
NO
NO
Fα1 = 1.E76
α1 = α Fα1 = FT
HF3 = F3
CH3(I) = C3(I)
CHT(I) = CT(I)
BEGIN DETERMINATIØN
FUNCTIØN
VALUE 2ND MIN.
NMIN = 4 MFLAG = 0
500
NO
NMIN = 4
YES
NIT = NIT + 1
NIT = 0
NO
Fα1 = 1.E76
YES
YES
NO
FT < Fα1
NO
72
YES
NIT = 2
NO
YES
α = α1 FT = Fα1
F3 = HF3
C3(I) = CH3(I)
CT(I) = CHT(I)
NEXT PRØBLEM
NO
FT < F3
YES
C(I) = CT(I)
C(I) = C3(I)


α
< 0
> 0
= 0
ALPHA IS NEGATIVE
NEXT PROBLEM
MK = MK - 1
$|x_3 D^{(k)} + \alpha D^k| \leq 0.01$
YES
519
NO
α < EPS
YES
225
NO
MK = 0
YES
519
NO
CALL EVAL
(α)
(FT, CT)
$|g^{(k+1)} H^{(k)} g^{(k)} \leq 0.001 \sum_{ij} g_i^{(k+1)} H_{ij}^{(k)} g_j^{(k)}|$
YES
519
NO
DT = $g^{(k+1)} H^{(k)} g^{(k)} > 0$
YES
X3 < α
YES
NO
NO
X3 > α
X2 = X3
F2 = F3
X3 = α
F3 = FT
C3(I) = CT(I)
NO
YES
X1 = X3
F1 = F3
X3 = α
F3 = FT
C3(I) = CT(I)
X1 = α
F1 = FT
X3 = α
F3 = FT
501


519
CALL EVAL
(α)
(FT,CT)
FT > FN
YES
25
NO
502
NMIN = 2
YES
NIT = 0
YES
NO
Fα1 = 1.E76
α1 = α Fα1 = FT
HF3 = F3
CH3(I) = C3(I)
CHT(I) = CT(I)
BEGIN DETERMINATIØN
FUNCTIØN
VALUE 2ND MIN.
NMIN = 4 MFLAG = 0
500
NO
NMIN = 4
YES
NIT = NIT + 1
NIT = 0
NO
Fα1 = 1.E76
YES
YES
NO
FT < Fα1
NO
72
YES
NIT = 2
NO
YES
α = α1 FT = Fα1
F3 = HF3
C3(I) = CH3(I)
CT(I) = CHT(I)
NEXT PRØBLEM
NO
FT < F3
YES
C(I) = CT(I)
C(I) = C3(I)


CALL
HUP
$(H^{(k)})$
$(H^{(k+1)})$
FN = FT
$X^{(k+1)}=X^{(k)}-\alpha H^{(k)}g^{(k)}$
$D^{(k+1)}=-\alpha H^{(k)}g^{(k)}$
PRINT NEW
PARAMETER
VALUES
$|(D_i^{(k+1)})| \leq EPS|x_i^{(k+1)}+0.001|$
$|(g_i^{(k+1)})| \leq EPS|x_i^{(k+1)}+0.001|$
YES
NO
ITER = ITER + 1
g(I) = C(I)
3
YES
ITER < MAX
NO
END
END
ABS. CONVERGENCE
SATISFIED
MAX. NUMBER OF
ITERATIONS
EXCEEDED


25
MFLAG = 0
YES
NO
RESTART PRØBLEM
WITH
H = I MK = 30
MFLAG = 1
5
NIT = 1
YES
NEXT PRØBLEM
NO
NIT = NIT + 1
GRADIENT
SEARCH
HAS
FAILED
502
225
YES
MFLAG = 0
NO
GRADIENT
SEARCH
HAS
FAILED
NO
NMIN = 2
YES
BEGIN DETERMINATIØN
ØF SECØND MINIMUM
500
NO
NMIN=4
Fal < FN
YES
END
72

8. References.

[1] Barnes, J.G.P. "An algorithm for solving non-linear equations based on the secand method". The Computer Journal, Vol. 8, April 1965, nr 1, pp. 66-72.

[2] Broyden, C.G. "A class of methods for solving non-linear simultaneous equations". Mathematics of Computation, Vol. 19, 1965, pp. 577-593.

[3] Broyden, C.G. "Quasi-Newton methods and their application to function minimization" Mathematics of Computation, Vol. 21, July 1967, nr 99, pp. 368-381.

[3a] Curry, H.B "The method of steepest descent for non-linear minimization problems". Quart. Appl. Math. Vol 2, 1944, pp. 258-261.

[3b] Davidon, W.C. Variable metric method for minimization. A.E.C. Research and Development Report, ANL-5990, November 1959, 21 pages.

[4] Dieben, L. Enkele problemen in de economie -de vraag naar productiefactoren-methoden bij niet-lineaire schatting. Term of probation report of the Tilburg School of Economics, Social Sciences and Law, 1971.

[5] Fletcher, R. and M.J.D. Powell "A rapid convergent descent method for minimization". The Computer Journal, Vol. 6, July 1963,pp. 163-168.

[6] Goldfarb, D. A family of variable metric methods derived by variational means. Working paper (to be published in Math. of Comp.)

[7] Hartley, H.O. "The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares". Technometrics, Vol. 3, May 1961, nr. 2, pp. 269-280.

[8] Heuts, R.M.J. and W.H. Vandaele
Numerical results of quasi-Newton methods for unconstrained function minimization.
Research memorandum 31, Tilburg Institute of Economics, June 1971.

[9] Leon, A.
A comparison among eight known optimization procedures.
Internal working paper, no 20, Space Sciences Laboratory, University of California, Berkely, August, 1964.

[10] Macon, N.
Numerical analysis.
John Wiley & Sons, Inc., New York, 1963.

[11] Marquardt, D.W.
"An algorithm for least-squares estimations of non-linear parameters".
J.Soc. Indust. and Appl. Math., Vol. 11, 1963, nr 2, pp. 431-441.

[12] Oliver, F.R.
"Another generalization of the logistic growth function".
Biometrika, Vol. 37, January 1969, nr 1, pp. 144-147.

[13] Powell, M.J.D.
On the convergence of the variable metric algorithm.
Report nr. T.P. 382, A.E.R.E., Harwell, October, 1969.

[14] Rao, C.R.
Linear statistical inference and its applications.
John Wiley & Sons, Inc., New York, 1965.

[15] Rosen, E.M.
A review of quasi-Newton methods in non-linear equation solving and unconstrained optimization.
National conference of the A.C.M. proceedings of the 21-st conference.
Washington D.C., Thompson Book Co., 1966, pp. 37-41.

[16] Shanno, D.F.
Conditioning of quasi -Newton methods for function minimization.
Center for mathematical studies in business and economics, University of Chicago, Report 6910 (revised), August 1969, pp. 20.
(Submitted to Math. of Comp.)

[17] Shanno, D.F. and P.C. Kettler — Optimal conditioning of quasi-Newton methods. Center for mathematical studies in business and economics, University of Chicago, Report 6937, August 1969, p. 16.

[18] Timmermans, A.J.M. — Een vergelijkend numeriek onderzoek van vier methoden voor het minimaliseren van sommen van kwadraten van niet-lineaire functies. Term of probation report of the Technische Hogeschool Eindhoven, 1971.

[19] Weeg, G.P. and G.B. Reed. — Introduction to numerical analysis, Waltham, Blaisdell Publishing Company, 1966.

A.K.

PREVIOUS NUMBERS:

EIT 1	J. Kriens *)	Het verdelen van steekproeven over subpopulaties bij accountantscontroles.
EIT 2	J. P. C. Kleynen *)	Een toepassing van „Importance sampling".
EIT 3	S. R. Chowdhury and W. Vandaele *)	A bayesian analysis of heteroscedasticity in regression models.
EIT 4	Prof. drs. J. Kriens *)	De besliskunde en haar toepassingen.
EIT 5	Prof. dr. C. F. Scheffer *) . . .	Winstkapitalisatie versus dividendkapitalisatie bij het waarderen van aandelen.
EIT 6	S. R. Chowdhury *)	A bayesian approach in multiple regression analysis with inequality constraints.
EIT 7	P. A. Verheyen *)	Investeren en onzekerheid.
EIT 8	R. M. J. Heuts en Walter H. Vandaele	Problemen rond niet-lineaire regressie.
EIT 9	S. R. Chowdhury *)	Bayesian analysis in linear regression with different priors.
EIT 10	A. J. van Reeken	The effect of truncation in statistical computation.
EIT 11	W. H. Vandaele and S. R. Chowdhury *)	A revised method of scoring.
EIT 12	J. de Blok	Reclame-uitgaven in Nederland.
EIT 13	Walter H. Vandaele	Mødscø, a computer programm for the revised method of scoring.
EIT 14	J. Plasmans *)	Alternative production models. (Some empirical relevance for postwar Belgian Economy)
EIT 15	D. Neeleman	Multiple regression and serially correlated errors.
EIT 16	H. N. Weddepohl	Vector representation of majority voting.
EIT 17	Walter H. Vandaele	Zellner's seemingly unrelated regression equation estimators: a survey.
EIT 18	J. Plasmans *)	The general linear seemingly unrelated regression problem. I. Models and Inference.
EIT 19	J. Plasmans and R. Van Straelen .	The general linear seemingly unrelated regression problem. II. Feasible statistical estimation and an application.

EIT 20	Pieter H. M. Ruys	A procedure for an economy with collective goods only.
EIT 21	D. Neeleman *)	An alternative derivation of the k-class estimators.
EIT 22	R. M. J. Heuts	Parameter estimation in the exponential distribution, confidence intervals and a monte carlo study for some goodness of fit tests.
EIT 23	D. Neeleman	The classical multivariate regression model with singular covariance matrix.
EIT 24	R. Stobberingh	The derivation of the optimal Karhunen-Loève coordinate functions.
EIT 25	Th. van de Klundert	Produktie, kapitaal en interest.
EIT 26	Th. van de Klundert	Labour values and international trade; a reformulation of the theory of A. Emmanuel.
EIT 27	R. M. J. Heuts	Schattingen van parameters in de gamma-verdeling en een onderzoek naar de kwaliteit van een drietal schattingsmethoden met behulp van Monte Carlo-methoden.
EIT 28	A. van Schaik	A note on the reproduction of fixed capital in two-good techniques.
EIT 29	H. N. Weddepohl	Vector representation of majority voting; a revised paper.
EIT 30	H. N. Weddepohl	Duality and Equilibrium.
EIT 31	R. M. J. Heuts and W. H. Vandaele .	Numerical results of quasi-newton methods for unconstrained function minimization.
EIT 32	Pieter H. M. Ruys	On the existence of an equilibrium for an economy with public goods only.
EIT 33	. . . . .	Het rekencentrum bij het hoger onderwijs.

EIT 1972

*) not available